

How to help universities meet obligation and opportunity

A landmark report on British higher education highlights an undercapitalized system ever more driven by a growing diversity of customers and the breakdown of trust between the academic community and the state.

The arrival of mass higher education in Britain — as in other industrialized countries — has been a mixed blessing for many of the institutions concerned. The desire that a continuously increasing proportion of the population should be able to benefit from a university course is unexceptional. But expansion has brought growing pressures, particularly as budgetary resources have not kept up either with the expanding teaching load or with the investment required to sustain across-the-board excellence in research.

As the National Committee of Inquiry into Higher Education, chaired by Sir Ron Dearing, reported last week (see page 413), there are serious problems that need to be addressed if universities are properly to fulfil the new roles and responsibilities that have been placed on them. Some will require more money, and it was inevitable that the government's main response to the Dearing committee's report focused on its decision to end an era of free university tuition. Others, however, will require cultural and attitudinal changes in universities themselves, and it is on the government's ability to stimulate and secure these changes without major disruption that its success in implementing Dearing's recommendations will be judged.

At the heart of these changes is the question of trust. 'Old style' higher education was based on an unwritten contract between universities and the state, at the core of which was the concept of academic freedom — the idea that universities should be entirely free to choose what to teach and what to research. Appropriately, this purity has been abandoned, but the trust implicit in this contract has also disappeared under increasing demands for demonstrable value for money. No longer is it accepted that what is good for universities is good for the country.

This breakdown of trust has had its own unfortunate consequences. Some are manifest in excessive demands for accountability from the research community, resulting in hours spent filling in forms that could be more productively spent at the laboratory bench. In turn, many researchers feel under pressure to base their research and fund-seeking strategies on maximizing personal, relatively short-term gain, regardless of whether this is in the long-term interests of their institutions, or indeed of those they work with (reflected, for example, in the use of funds allocated for overheads to pay for short-term research assistants with little reflection on their longer-term career prospects).

Can this trust be rebuilt? Four years ago, the government sought in a white paper (policy document) to redefine the missions of the research councils in terms of an explicit commitment to wealth creation and the quality of life. Many of Dearing's complementary recommendations for universities are expected to be taken up in a new white paper on higher education due in the autumn. Both initiatives reflect pressure for structural changes in the relationship between the research community and the state. But both will be undermined if a proper balance is not restored between freedom and accountability.

How can a proper partnership be re-established? The government can display its commitment by providing the financial support essential for maintaining the health and vitality of the science base. It is dangerous to use figures of current performance, based on previous invest-

ment, to deflect charges of underfunding. The real value of such figures lies in demonstrating the cost-effectiveness of British science — and thus the potential gains to be had from new investment.

On the scientists' side, there is a need to combine a commitment to scientific excellence (judged now by international standards) with clear-headed strategic thinking. On the first, Dearing's suggestion that foreign researchers be invited to participate in the Research Assessment Exercise is to be welcomed, provided that their input is tempered with an acknowledgement of the other dimensions of the value of research.

As to the second, part of the value of instigating a more rigorous approach to the refunding of indirect costs is that it will encourage universities to check more robustly where this money does in fact go. The Dearing report itself emphasizes the need to ensure that the temptation to use overhead costs to fund research rather than infrastructure is "strenuously avoided". Medical charities' reluctance to cover indirect costs is based not on principle, but on the inability of universities to demonstrate that the money is not being used to replace the chapel roof.

There are inevitably prices to be paid for moving towards both greater selectivity and more strategic thinking. An excessively rigorous commitment to excellence always risks creating difficulties for those at the bottom of the ladder. It is essential that any attempt to distinguish between 'teaching' and 'research' departments still enables individuals to move flexibly between the two roles. And there remains a lack of imaginative suggestions as to how teaching performance can be awarded the same credit as research achievement — indeed what universities can do to give teaching the status and rewards that it deserves.

Another danger lies in the extra administrative burden created by a combination of tighter cost control, increased public accountability, and even strategic planning itself. The additional work-load must not rest excessively on researchers whose strongest skills and talents lie elsewhere. Funding bodies must be convinced that the extra administrative costs are themselves a legitimate overhead. It is surely no coincidence that those US universities that charge the highest overheads are also the most scientifically productive.

Finally, industry must not be let off the hook. There is a clear logic in the insistence of industrial leaders that it is the government's responsibility — not theirs — to ensure the health of the nation's science base. It is also clear that, where self-interest is high, industry will remain keen to build links with groups that meet its immediate needs. More difficult, but no less vital, is private sector support for generic initiatives whose benefits lie between the two; here, a whole-hearted commitment to initiatives such as Dearing's proposed Industrial Partnership Development Fund would be a welcome sign of good faith.

The main message of the Dearing report is the need for a new 'compact' between the different stakeholders with an interest in the continued health of British universities. If the language is uncomfortable to some, this is a measure of its novelty rather than its inappropriateness. But rebuilding both trust and commitment, on all sides, is the real, if at times painful, task facing the UK government as it works out how to respond. □



Clinton pulls out the stops in bid to win backing for carbon cuts

[WASHINGTON] President Bill Clinton met prominent scientists at the White House last week and began what he promised would be a sustained campaign to win US public support for action on climate change in the run-up to the international meeting in Kyoto, Japan, in December.

"There is ample evidence that human action is already disrupting the climate," Clinton said. "We can see the train coming but most ordinary Americans, in their day-to-day lives, can't hear the whistle blowing." This is true throughout the world, Clinton said, and represents "a great test for our democracy".

Clinton and Vice-President Al Gore met seven scientists, including Nobel prizewinners Sherwood Rowland, Mario Molina and Henry Kendall, before banks of television cameras and an audience of officials. The scientists gave presentations on different aspects of climate change, and were questioned by Clinton and Gore.

Clinton asked John Holdren, a physicist and professor of public policy at Harvard University, "why people are underrating the importance of the issue" and what he, as president, should do about it.

In a swipe at global warming sceptics, Holdren responded that there is "a solid consensus among scientists who have studied these matters seriously". He asked for early action to cut carbon emissions because "every year of business as usual, the more difficult it becomes to do anything about it".

Clinton said after the meeting that it had been the beginning of "a consistent, long-term effort" to educate the public on the issue. "I believe it will give us the support we need to take the action that we are going to take," he said in remarks directed toward the Kyoto meeting, at which countries are due to agree



Top class: Vice president Al Gore and president Bill Clinton listen as Mario Molina (right) of MIT explains the dynamics of climate change at a heavily publicized seminar in the White House.

on moves to curb carbon emissions.

The White House event appeared to indicate that Clinton has been convinced by Gore and other environmentalists in his administration to try to build public support for action on climate change, as the environmentalists have been seeking.

But it remains to be seen how far he will take the effort in the face of public apathy on the issue and widespread hostility to taxation or any other government action aimed at reducing energy consumption.

The event received modest media coverage, and the following day the Senate unanimously backed a resolution from Senators Robert Byrd (Democrat, West Virginia) and Chuck Hagel (Republican, Nebraska), calling on the administration not to agree on anything in Kyoto that would damage the US economy or set emission limits for developed countries while leaving develop-

ing countries without restrictions.

The Senate will have to ratify any agreement signed at Kyoto, and the Byrd resolution is seen as a warning that it will not accept an agreement that places most of the burden for reducing emissions on the United States.

Having failed in earlier efforts to modify the Byrd resolution, environmentalist senators and administration officials now say they can live with it. "The notion that developing countries need to be part of the solution is something we agree with," said Katie McGinty, head of the environment office in the White House.

McGinty says that the United States will call for the creation of a new category of countries at Kyoto which will take on binding obligations under the climate treaty as they reach a certain level of economic development. An automatic graduation process would transfer developing countries into the new category as their economies developed, she adds.

Earlier in the week, Gore had re-emphasized his strong commitment to action on climate change. "Will we accept responsibility for protecting our children from what we're doing or will we pretend it doesn't matter?" he asked an audience of engineers from the Partnership for a New Generation of Vehicles, a programme to develop a 'clean' car. "I think the question answers itself."

Jack Gibbons, the president's science adviser, told reporters that a carbon tax "should be considered" in response to global warming, along with other options. Asked about the preference of US consumers for large, uneconomical 'utility vehicles', Gibbons said: "We have to learn to better understand the long-term implications of our short-term decisions."

Colin Macilwain

IPCC reports 'should be more user friendly'

[LONDON] Members of the United Nations climate convention will this week meet in Bonn to consider the structure and content of the third five-yearly assessment of the world's climate.

The members – from the convention's Subsidiary Body on Scientific and Technical Advice – will discuss the assessment at a meeting to be conducted by the Intergovernmental Panel on Climate Change (IPCC) under

its new chair, Bob Watson.

Some countries have indicated they would like the new report to reflect better the different views of scientists and not to shy away from delving into debates on climate change, areas of uncertainty, and 'alternative' views. Environmentalist groups, too, support this direction, according to a report in the newsletter *eco*.

The second assessment

report, published last year, contained the famous phrase: "the balance of evidence suggests a discernible human influence on global climate", which convinced governments of the need to take action.

Switzerland has said it is concerned that this report was not user-friendly enough – particularly the synthesis, which condensed the three-volume document into a summary for policy-makers. □

NASA urged to extend all solar satellites

[WASHINGTON] The US space agency NASA should keep all its current Sun-watching satellites operating until the end of 2001, and should expand the community of researchers analysing their data. That is the conclusion of a panel of experts from outside the agency who had been asked by NASA to review its solar-terrestrial physics programme.

NASA will begin running out of money next year to operate the satellites, and had asked the 'senior review' panel to help it to decide which were most valuable scientifically (see *Nature* 387, 747; 1997).

But, rather than pick winners and losers, the panel recommended that all eight missions under consideration should have their lives extended. The satellites, which include the European-American Solar and Heliospheric Observatory (SOHO), Ulysses, WIND, POLAR and Voyager, study Sun-Earth interactions from different vantage points in the inner Solar System. They were launched when the Sun was relatively quiet. But now, says the panel, "an opportunity exists, not to be repeated for generations, to use this same array of spacecraft... during the onset and peak of solar maximum conditions".

In fact, says the panel, "it can be argued that the massive [\$3 billion] investment in these spacecraft will have been worthwhile only if they are used to provide the same level

of information on the important processes of solar maximum", which begins shortly after the turn of the century.

While it found "no redundancy or duplication" among the missions, the panel did say that savings could be achieved in spacecraft operations. And existing scientific teams for the satellites could be pared down by reducing the number of co-investigators. Some 350 scientists are active in the programme, and that number could come down to about 200 in a minimum level 'Solar Maximum Campaign'.

But that will not be enough, says the panel, because understanding the solar maximum is a trickier scientific problem than understanding the quiet Sun. So they "strongly recommend" that NASA should fund an interdisciplinary investigator programme at \$15 million a year, beginning in 1999, to make up for the shortfall in analytical manpower. Principal investigators would receive additional funds to prepare the satellite data, but the interdisciplinary programme funding would go to support about 125 scientists not directly involved with the project, who would compete for grants.

The panel, chaired by Leonard Fisk of the University of Michigan, a former head of NASA space science, estimates the total funding required for a vigorous Solar Maximum Campaign at \$26 million between 1998 and

2001. "It is at this level, and only this level, that the review panel is prepared to guarantee that an effective program to solve the most complicated, but also the most significant, problems in Sun-Earth-Heliosphere connections will succeed," the panel says.

Whether the group's strong stand will translate into more money for NASA remains to be seen. Guenter Riegler of the agency's Office of Space Science says the recommendations represent "the ideal best of all worlds". Right now, he says, the agency expects to receive only between 60 and 70 per cent of the money it would need to pay for such a programme. The rest presumably would have to come from an additional funding request to Congress beginning with fiscal year 1999.

The review panel acknowledges that the space agency may not be able to afford the programme it recommends. But, rather than start turning off satellites or cutting back on science immediately to make ends meet, the group recommends that NASA should continue a vigorous programme for as long as the money holds out.

"An extraordinary opportunity will be lost if the full program is not pursued through the expected solar maximum in 2001," the panel says. "However, it would be a greater tragedy not to use fully the available assets while they can be supported."

Tony Reichhardt

Three weeks into mission, Sojourner dusts off Mars' secrets



[LONDON] Mars — at least in the small patch of the Ares Vallis being explored by the US Pathfinder mission — is both more varied and more

uniform in appearance than expected.

Images from the optical camera on the mission's lander have revealed a relatively colourful landscape, with contrasting reds and greys; yet the elemental compositions of different soils and rocks measured by the Sojourner rover are mostly very similar.

Pathfinder is still operating well, after more than three weeks on the Martian surface. Despite communications problems which have slowed the rover's progress — for example, giving it the weekend off on 19 and 20 July — several rocks and soil samples have been analysed using its alpha proton X-ray spectrometer (APXS).

The first rock to be analysed, Barnacle Bill, is the most intriguing so far, being similar in composition to andesites, or basaltic andesites, on Earth. The rock's relatively high silica content suggests it

may have been reheated or heated in a watery environment.

In contrast, Yogi — a larger rock that Sojourner took a few days to get to grips with properly — seems to contain lower levels of silica. It may have an ordinary basaltic composition, as expected for most of Mars's crust; but it is covered with dust, which may have made the measurements useless. "Next time, we should take a clothes-brush", comments Rudy Rieder, principal investigator on the APXS.

As a result of this experience, the Pathfinder team decided to clean the next rock. After an initial APXS measurement of the pale-looking Scooby Doo, they rolled the rover back and forth across it, spinning the wheels to try to scrape off as much soil as possible. (Similarly pale material was dug up by Sojourner's wheels on one occasion and may be widespread just below the surface.)

This exercise made little difference to the appearance of Scooby Doo, so it is probably a white rock, rather than a dark one covered in pale dust.

Before the scraping, Scooby Doo appeared to have the same composition as the soils. "That's surprising, as it is a very

different colour," says co-investigator Johannes Brückner. "There might be some salt there, but we need more analysis to be able to tell."

The diversity of colour of soil patches may be due to differences in particle size, varying degrees of weathering, or chips of rock scattered in darker soils. Surprisingly, differences in elemental composition do not seem to be responsible.

Martian soil is divided into two grades: fine, bright-red 'drift', with particles smaller than about 40 µm; and coarser, deeper-red 'lag'. Drift is carried around the planet by dust storms, becomes thoroughly mixed on a global scale and therefore probably has an average crustal composition. Lag, on the other hand, had been thought to be of more local origin and thus of varied composition.

But Sojourner has not been able to detect any difference, finding identical compositions to the Viking lander soils in every case. One suggestion is that Lag is compacted drift and the white rocks, such as Scooby Doo, may be compacted dust too, although they must have undergone some kind of chemical or mineralogical change if they are

Stephen Battersby

Japan seeks funding for new drilling ship

[TOKYO] Earth scientists and deep-sea drilling engineers from 17 countries, meeting in Tokyo last week, unanimously endorsed a Japanese plan to build a new deep-sea drilling ship. The ship is intended to play a key role in the future of the international Ocean Drilling Program (ODP) beyond 2003, when the current phase of the programme ends.

The proposed ship, part of a Japanese initiative called 'Ocean Drilling in the 21st Century', would adapt drilling technology from the petroleum industry to allow drilling to much greater depths in the Earth's crust and in more difficult geological terrain than is possible with the current programme's drilling ship, the *JOIDES Resolution*.

In particular, the new ship would use 'riser technology', in which the drilling stem is encased in a rigid pipe that would rise off the sea floor in water depths of up to 2,500 metres and extend up to the ship. The riser pipe would allow the circulation of drilling mud to facilitate drilling to depths of up to 7 km beneath the sea bed — several times the maximum depth achieved so far in the ODP.

The ship, first proposed by Japan's Science and Technology Agency (STA) in 1990 (see *Nature* 346, 307; 1990), has become critical to the ODP's future. Last year, it was effectively incorporated into the programme's long-term plan, as this calls for a two-ship programme beyond 2003 with a riser-equipped drilling ship and another ship similar to the *JOIDES Resolution*.

But it remains to be seen whether the ODP partners, which include the United States, Japan, France, Germany, Russia, Canada

and the United Kingdom, will be prepared to provide the extra funds needed to run the ambitious programme — even if the Japanese ship, which is expected to cost about US\$500 million, gets budgetary approval.

The Tokyo meeting, called the Conference on Cooperative Ocean Riser Drilling (CONCORD), was timed to give maximum support to STA's attempts to win money for the ship in the agency's budget request for the next fiscal year, due to be drawn up at the end of August. But according to Hajime Kinoshita, director of the deep-sea research department of the Japan Marine Science and Technology Center, which would operate the ship, the agency has "more or less given up" hoping for a budget for the ship next year because of the government's mood of fiscal restraint.

Instead, it will lobby for a budget of about US\$35 million to build drilling equipment for the ship in the hope that this will lead to money to design the ship in the following fiscal year. That means the ship cannot be completed until 2003 at the earliest.

The two-ship programme will cost about US\$120 million a year to run, nearly three times ODP's current annual budget of US\$45 million, more than half of which comes from the US National Science Foundation. Last year, France and the United Kingdom considered withdrawing from the programme because some felt it was not giving sufficient value for money and that much ODP research was ending up in 'grey literature' in ship reports (see *Nature* 379, 193; 1996). But according to Kiyoshi Suyehiro of the Ocean Research Institute of Tokyo University, one of CONCORD's organizers, France and the



The next generation of drilling ship will drill deeper than *JOIDES Resolution* (above).

United Kingdom are now fully committed to staying with the ODP until at least 2003.

Xavier Le Pichon, of France's Laboratoire de Géologie of the Ecole Normale Supérieure, said in a keynote speech that though there had been "strong criticism" of the ODP, including some by himself, the ocean drilling community is making "major efforts to change".

Jim Briden, director of the Environmental Change Unit at the University of Oxford, and the UK representative on the ODP executive, said that the United Kingdom "fully supports" expanding the programme. He acknowledged that getting the required funds would be a "challenge", but was confident that six years is long enough in which to do this.

Briden says that two additional sources of funds have been identified: industry and the European Commission. "There is nothing in current EC research programmes that matches the ODP's work," he said. He added that negotiations had begun with Europe's oil industry, which is interested in the new riser drilling technology.

Le Pichon stressed the need to involve a wider community of scientists in the ODP's future, including those involved in the International Continental Scientific Drilling Program. Indeed, it was generally recognized by participants that more countries will have to join the ODP if it is to raise the funds for a two-ship programme. China, for example, was represented at the meeting, and is about to become an ODP member, but with a greatly reduced membership fee.

Japan is also looking to ODP partners to help with the costs of their proposed ship through, for example, the provision of ship-board laboratory and onshore facilities.

The meeting identified several scientific goals for the new ship, including the extension of studies of the climate variability record to 180 million years; the search for bacteria deep in marine sediments and the ocean crust; and drilling into the upper mantle through the ocean crust. But the meeting decided unanimously that drilling should initially concentrate on the active seismic region in the subduction zone along Japan's coast so as to gain a better understanding of earthquakes through direct long-term observation.

David Swinbanks

German truce on genome data access

[MUNICH] Germany's pharmaceutical industry has agreed to forgo an offer of three months' privileged access to genome sequence data generated by the German Human Genome Project.

Some American and British researchers had threatened to exclude German colleagues from international collaborations in human genome sequencing research, and to block their access to international sequence databases, if the privileged access was maintained.

Last year the Verein zur Förderung der Humangenomforschung (Förderverein), an industrial funding association of eight major German drug companies, agreed with the research ministry that member companies should have privileged access to data. This generated sharp criticism in the international research community, as other countries do not grant their industries a similar privilege (see *Nature* 387, 536; 1997).

In a letter to the research ministry, the Förderverein says that both genomic and cDNA sequence information will be covered by the waiver of its privileged access. "This is a relatively small price to pay for industry", says Werner Schiebler, director of New Technologies Licensing at Hoechst AG in Frankfurt am Main, a member company of the Förderverein. "It should help to build confidence between pharmaceutical companies and academic scientists, encouraging the latter to make use of patents and licensing agencies."

The waiver is likely to resolve the conflict over the rules of Germany's Human Genome Project. But because differences between European and US patent laws lie at the heart of the problem — in particular, the existence of a 'grace period' under US law for filing patents — the Förderverein says it sees "a clear need to harmonize international patent law".

Quirin Schiermeier

US warned about complacency on Japanese economy

[WASHINGTON] The United States should beware of economic competition from Japan and redouble its efforts to build links with that country's science and technology base, even at a time when US industry has regained its competitive edge, according to a report this week from the National Research Council (NRC).

The report notes that Japan continues to spend more on non-military research and development, and employ more scientists per head of population, than any other nation, despite its recent economic difficulties. It predicts that the Japanese government's new emphasis on basic scientific research will build a formidable base of research results which the United States could be ill-equipped to tap into.

"If the United States is complacent and underestimates Japan's ability to bounce back, we could find ourselves facing some of the same problems we faced in the early 1980s," says Eric Bloch of the Council on Competitiveness, who chairs the Committee on Japan at the NRC, the operating arm of the National Academy of Sciences.

Jim Martin, a research manager with Rockwell who chairs the task force that produced the report, says that "there is a feeling of complacency in every sector of America" about competition from Japan. But the report urges the US government to act on several fronts to "make the US-Japan science and technology relationship more balanced".

A new 'watch list' should be established, it says, to track the handling of patents applications by US citizens in Japan, and to ensure that patent protection is enforced. The US government should continue to expand its programmes to acquire important Japanese technical information, and make it available to US companies and scientists.

It should expand use of the US-Japan Science and Technology Agreement, an umbrella agreement which is supposed to simplify scientific collaboration, and develop new ways to measure its performance.

Despite some progress, Bloch says, the science and technology relationship between the two countries is "one of asymmetry", with far fewer US scientists and engineers visiting Japan than vice versa.

But, in the six years since Congress requested the report, the United States has enjoyed surging economic growth and has come to regard Japan as less of a threat, panel members say. Asked for the source of this complacency, Bloch singles out the US press. But he admits the NRC itself may soon close its Office of Japan Affairs.

Colin Macilwain

'Skewed medical goals' revealed by Indian survey

[NEW DELHI] A bibliometric study of medical research in India has concluded that much of the work being done is not directly relevant to the most urgent health needs of the population.

According to government statistics, diarrhoeal, respiratory, infectious and parasitic diseases account for most deaths and morbidity in India (see Table 1). But the study shows that researchers have been more active in studying diseases such as cancer and neurological disorders, whose significance is felt to be relatively limited, rather than more widespread diseases such as malaria, which affects 2.5 million Indians each year (see *Nature* 386, 536; 1997).

Subbaiah Arunachalam of the M. S. Swaminathan Research Foundation in Madras — now Chennai — carried out a scientometric study in 1995, based on Indian medical papers cited in Science Citation Index, which showed a similar result. But this covered only one Indian medical journal out of the 250 or so published.

He has now carried out a study based on data from Medline, which indexes 30 Indian medical journals. The results, published in the journal *Current Science* (72, 912; 1997) of the Bangalore-based Indian Academy of Sciences, reached the same conclusion.

Arunachalam found that, between November 1987 and December 1994, Indian authors published 18,224 articles in 45 medical fields in 1,368 journals. One conclusion was that, in terms of the number of papers published, neither tropical medicine nor respiratory diseases figure in the top 10 fields in Indian medical research (see table 2).

Indian researchers published 584 papers in 101 journals in neuroscience, 1,367 papers in 94 journals in pharmacology, and 821 papers in 56 journals on cancer. But they published only two papers in an epidemiology journal in seven years.

Although agricultural research played an important role in transforming India from a food-deficient country into one with food

Table 1 **Leading causes of mortality and morbidity in India 1991-93**

Mortality	Diarrhoeal diseases
	Respiratory diseases
	Infancy diseases
	Pneumonia
	Infectious and parasitic diseases
Morbidity	Respiratory diseases
	Diarrhoeal diseases
	Malaria
	Whooping cough/measles
	Neonatal tetanus

Source: WHO

surpluses, "medical research in India, but for a few exceptions, has not covered itself with glory despite the fact that medicine enjoys a better status and image than agriculture in Indian society," writes Arunachalam. He says the question of relevance is especially important in a developing country where scarce resources have to be used judiciously.

The Indian Council of Medical Research (ICMR) has challenged the study's conclusions, denying any mismatch between the work of its researchers and national needs. All the 21 ICMR institutes and five regional medical research centres in different parts of India "direct their efforts for research on diseases or disciplines which are on the national health agenda," the council said in a statement. "Evaluating their contributions in terms of mere publications in indexed (or even other) journals would be not only unfair but unrealistic."

ICMR's deputy director general, Lalit Kant, says that most Western databases, including Medline, cover diseases of the developing countries inadequately. "Any analysis of the relevance of medical research in India should be supplemented with authentic information from other databases like tropical disease bulletins and national databases," he says.

Marthanda S. Valiathan, a leading heart surgeon and vice-chancellor of the Manipal Academy of Higher Education, says that Arunachalam's findings "reveal a lopsided order of priorities in Indian medical research". Valiathan traces the origin of the mismatch to the nineteenth century, when Indians started using Western research tools and techniques without developing their own.

But Balasubramaniam Ramamurthi, one of India's leading neurosurgeons, based at the Voluntary Health Service Centre in Chennai, warns against blaming scientists. "Abolishing diarrhoea, tuberculosis and malaria requires public, political and administrative action, and not research," he argues.

K.S. Jayaraman

Table 2 **Indian research papers covered by Medline, 1987-94, by subfields (first 10)**

Subject	No. of journals	No. of papers
General medicine	57	2394
Paediatrics	43	1420
Pharmacology	94	1367
Immunology	74	928
Pathology	48	916
Oncology	56	821
Surgery	68	750
Cardiovascular	41	663
Gastro	26	606
Neuroscience	101	584

FDA seeks to close xenotransplant gap

[WASHINGTON] Ways to minimize the risks that transplanting animal organs into humans could trigger a new disease pandemic were discussed at a meeting to help build bridges between scientists and transplant surgeons last week.

The US government will revise its guidelines on xenotransplantation by the end of this year, partly on the basis of a dialogue between surgeons, who are eager to experiment with it, and virologists, some of whom fear that such transplants could transfer new infectious diseases to the human population, and perhaps trigger new pandemics.

The meeting was held in Bethesda, Maryland, by US government agencies to help establish such a dialogue. It heard that possible routes to reducing risk included the breeding of virus-free herds of donor animals, and the vaccination or antibody treatment of patients to eliminate virus from their blood after they receive a transplant.

Officials said that the two-day meeting had established a better understanding between surgeons and virologists. "We were quite pleased with the process," says Phil Noguchi, a senior official at the Food and Drug Administration (FDA), who helped to organize the meeting. "I'd say the virologists were somewhat less conservative than we might have expected."

But several virologists at the meeting said that the public health risks of xenotransplantation are very difficult to assess. Preston Marx of the Aaron Diamond AIDS Research Center in New York, who spoke about his studies of cross-species infection, said afterwards that the risks could be reduced but not eliminated.

Marx said that viruses have been trans-

ferred from animals to humans — by biting or blood-to-blood contact, for example — throughout history. But the result is usually a "dead-end infection", he says, as the virus is not transmitted between humans.

Experiments on the impact of transplants from pigs, the most likely animal donor species, to non-human primates could be useful. According to Ron Ferguson, a transplant surgeon at Ohio State University: "Primate trials with pig organs could well answer some of these questions and give some realism to policy, rather than having a policy based on fear of the unknown."

Draft guidelines on xenotransplants were published by the US Public Health Service last year, but some virologists say they are insufficiently restrictive. They point out that the transfer of viruses between species is inevitable: recent *in vitro* research at the Institute of Cancer Research in London, for example, demonstrated that a pig retrovirus can be infectiously transmitted to human cells (see *Nature Medicine* 3, 282–286; 1997).

But the actual risk that pig virus could then be transmitted between humans and cause disease is unknown. "We need to be creative in finding ways to define the risks in a form that the public can deal with," said Jay Fishman, a surgeon and researcher at the Massachusetts General Hospital in Boston. "If we don't collaborate openly, it will be very difficult to convince the public that we're not hiding something."

But some surgeons are impatient with scientists' efforts to assess risks that surgeons see as hypothetical. Ferguson reminded the meeting that transplantation was launched, as a medical practice, by doctors in London



hospitals during the Blitz early in the Second World War, and implied that it has been working over the objections of scientists ever since. "We have the history to sit down with another discipline, and mould something good," he said.

William Raub, deputy assistant secretary at the Department of Health and Human Services, promised that the meeting would help the department to prepare new guidelines "within the next several months" that would include rules for assessing the risks of each proposed research protocol.

About a dozen research protocols are under way in the United States, mostly involving pig cells and tissues and none involving human patients. Each has had to be approved individually by the FDA. The health department plans another workshop in the autumn, with patient representatives and regulators as well as scientists and surgeons, to discuss xenotransplant policy.

Colin Macilwain

University settles with patients over trade in 'stolen' embryos

[SAN FRANCISCO] The University of California has settled, at a cost of about \$14 million, 72 lawsuits filed by former patients of its fertility clinics at San Diego and Irvine, where two years ago it was alleged that eggs and embryos were stolen and implanted into other women or used for research.

Details of the agreement were not disclosed, under an order from the California Supreme Court to protect the privacy of the couples involved. According to the lawsuits, doctors at the clinics had given eggs to other couples without the consent of the women from whom they were taken (see *Nature* 376, 456; 1995 & 379, 756; 1996).

In addition, patients were included in research studies without their consent; drugs not approved by US regulators were sold to patients; and informed consent for medical procedures was at best sloppy and poorly documented. As a result, children were born

without the knowledge of the women who had produced their eggs.

The agreement must still be approved by Orange County Superior Court. James Holst, general counsel for the university, said in a statement that the university would continue to pursue settlement in the 30 cases remaining.

The Irvine Center for Reproductive Health is now closed, and the university's San Diego programme has reopened under new leadership, after being closed two years ago.

The three principal doctors in the case continue to face criminal charges and have been indicted on federal charges of mail fraud and income tax evasion. Two have left the country: Ricardo H. Asch is in Mexico and Jose P. Balmaceda is in Chile. Serio C. Stone is under house arrest in California. All have denied wrongdoing, and blame university officials for any paperwork mix-ups.

According to Laurel Wilkening, chancellor at the University of California at Irvine, the money for the settlement comes out of an insurance liability fund paid into by university medical centres. Tenet Healthcare Corporation, which now owns the hospital in which many of the cases originated, has agreed to help with the payment.

University officials say that the funds would not have been spent for any other purpose. But Gray Davis, California's lieutenant governor and an ex-officio member of the university Board of Regents, complained to reporters that taxpayers were once again paying to clean up a university management failure. He said managerial and fiscal controls must be improved.

In the wake of the scandal, the university set up task forces to look into supervision of research and quality of care at fertility centres.

Sally Lehrman

US and Europe may link up planned arrays

[WASHINGTON] European and US astronomers are considering merging their plans to build large millimetre-wave astronomy observatories, a move that could result in a single, more sophisticated facility.

But design differences between the US Millimeter Array (MMA) and Europe's proposed Large Southern Array (LSA) still have to be reconciled, and agreement reached on funding for the joint project, which would cost about \$400 million.

At a meeting last month at the headquarters of the National Radio Astronomy Observatory (NRAO) in Charlottesville, Virginia, both sides agreed that construction and operating costs should be split evenly between Europe and the United States. They also agreed that the array would be located at the 5,000-metre-high Chajnantor site in Chile's Atacama Desert, which had already been chosen for the MMA.

The US plan calls for 40 eight-metre antenna dishes, working together as an interferometer. The European concept is to have fewer dishes but to make them twice as large, so that the LSA's total collecting area would be five times that of the MMA.

The European observatory would not observe at submillimetre wavelengths, but would be better suited to observing very distant, faint objects such as galaxies in the early stages of formation. The MMA in contrast would view larger, more extended sources at higher frequencies.

Scientifically, the two sides should have lit-

tle problem agreeing on a common approach, says Robert Brown of the NRAO. But they will need to choose between two design strategies — either keeping the current, different-sized antennas, or coming up with a common, intermediate size.

Each side has something to gain from a merger, say those involved in the discussions. The US National Science Foundation (NSF), which is funding the MMA, has been pressing the NRAO to find an international partner. Talks with Japan, which is also proposing a large millimetre-wave array, have yielded no agreement.

The LSA could benefit from attaching itself to the more mature MMA project, which was included in next year's NSF budget request after more than 15 years of planning. The LSA will not even come up for funding until 2000, says Dennis Downes, an astronomer at the Institut de Radio Astronomie Millimétrique in France, who is participating in the merger discussions.

A joint venture would also bring down the LSA's cost of \$350 million, if Europe had to contribute only \$200 million as its part of the partnership.

Another advantage of teaming up would be sharing the cost of building roads, buildings and other infrastructure at the site in Chile. That alone could save some \$50 million, according to Downes. Brown says the NRAO is trying to encourage Japan to build in the same area, if it goes ahead with its array.

US funding for MMA — a proposed \$9



The proposed US Millimeter Array (above) has been under pressure to seek a partner.

million in 1998 — so far covers only design studies and prototyping. The NRAO plans to build working models of the system's antenna, receiver and other key components by 2000. If these are approved, the full system could be operating by 2006.

But supporters of the MMA got a rude shock two weeks ago, when the Senate appropriations committee that oversees the NSF's budget raided \$4 million of the MMA's \$9 million request to finance the Gemini Telescope, a higher priority, nearer-term project. Brown says the reduction, if it stands, would do "a lot of harm" to a project that already has its hands full developing cutting-edge technology on a short deadline.

He hopes the money will be restored when the House of Representatives and Senate reconcile their two spending bills in the autumn. "To throw out this challenge and then not give us the resources pulls the rug out from under us," he says.

Tony Reichhardt

Austria to go private in novel solution to grant distribution

[MUNICH] A report commissioned by the Austrian government has concluded that all government funding for industrial and academic research should be placed in the hands of a private organization. The new body would also develop long-term research strategies. The government approved the report in principle last month.

The report's authors — Albert Hochleitner, director general of Siemens Austria, and Arnold Schmidt, president of Fonds zur Förderung der Wissenschaftlichen Forschung (FWF), the funding agency for university research — had been asked to propose how industrial and academic research could be brought closer together. They recommend the establishment of an Office for Research and Technology (BFT), a state-owned private company that would distribute research grants to projects with a defined industrial relevance. The BFT would handle all existing funds, such as the ASch700 million (US\$54 million) of the FWF.

The report also suggests that the BFT would handle the funds of a new foundation,

KIR, which would support the creation of so-called centres of competence, with a budget around twice that of the FWF. Universities and industry would have to apply jointly for support for such new institutes.

The BFT would distribute a total of ASch4–5 billion in grants each year to researchers and universities. It would be scientifically advised and financially controlled by a Council for Research and Technology made up of the ministers for science and for economic affairs, four academics and four industrialists.

The report foresees that direct funding for Austria's universities, where 95 per cent of the country's research is carried out, would be gradually reduced over the next decade, and university scientists would instead compete for funds distributed by the BFT.

Despite the general approval of the plan by the Austrian government, the details of the proposal are likely to be changed by the time a formal cabinet decision is required in September. An official from the research ministry says that there are serious doubts

about whether entrusting a private company with political decisions of public interest would accord with Austria's constitution.

The research minister, Casper Einem, suggests that the BFT should be created as an endowment financed jointly by the research budgets of the ministries for research and economy, and proceeds from the privatization of state-owned industries. "The experts were free to think, but now it is up to us to reduce the concept to what is feasible," he says.

Some university staff strongly oppose reducing the research minister's influence on long-term science and technology strategies. They fear that grants may come to depend on the vote of a small but powerful panel of experts driven by economic interests.

"We do not support the concept of transferring responsibility for science policy [from the state] to bolster the power of a group of a few so-called experts," says Margit Sturm, secretary general of the Conference of Scientific University Staff. **Quirin Schiermeier**

UK universities told: 'think strategically'

David Dickson

Last week's long-awaited Dearing report has urged the UK government to take immediate steps to remedy the underfunding of university research. But at a price: universities, it says, must agree to do things differently.

The broadest inquiry into British higher education for more than 30 years has confirmed the conclusions — at least as far as research is concerned — of several recent reports: that university research is seriously underfunded, but that increased support should be provided only if universities are prepared to be more selective in their research strategies.

These are the two principal messages to emerge from the report published last week by the National Committee of Inquiry into Higher Education, set up last year by the then Conservative government. The committee was headed by Sir Ron Dearing, a former chairman of the Post Office.

The report's most controversial recommendations focus on its proposal that students should be required, for the first time for many years, to pay towards their university tuition. But the committee was also asked to look at the organization and funding of university research.

The education secretary, David Blunkett, appears to have accepted at least part of the panel's message. Answering a question in the House of Commons last week from Ian Taylor, the former Conservative science minister, Blunkett admitted that there is "a crisis in terms of equipment and the capacity to undertake research [in universities], and in terms of the buildings".

How far the government is prepared to go in remedying this situation will not be known until the publication of a promised white paper (policy document) on 'lifelong learning', due in the autumn.

Many of the Dearing committee's recommendations on research were initially drafted by a subcommittee chaired by Sir Richard Sykes, the chairman of Glaxo Wellcome, Britain's largest industrial investor in research. Members of the subcommittee included Sir Ron Oxburgh, rector of Imperial College, London, and other prominent academic and industrial scientists.

The committee points out that there is not an acute crisis in university research. Even though funding is tight, scientific output remains high. Indeed, it quotes the conclusion of a recent study by Sir Robert May, the government's chief scientist, that the standing of British research "remains competitive" in international terms.

But it also points out that many current strengths of British science are due to prior investment, and that this has not been maintained over the past decade. "Recently the UK has seen a decline in its share of world publications and citations which indicates that it may be falling behind," it says.

Of particular concern is the state of scientific equipment and the general scientific infrastructure, which has recently failed to receive adequate financing — partly because investment in equipment has been one of the easiest areas to cut under budget pressure. "The state of the research infrastructure in many universities is pitiful," says Sykes.

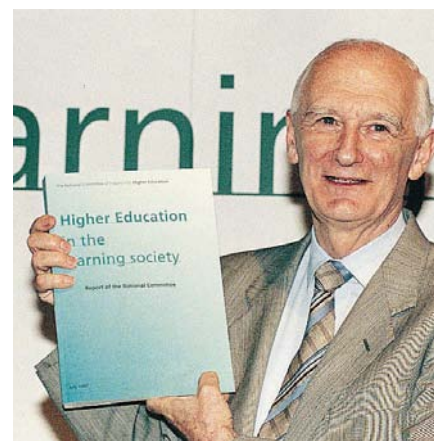
The lack of adequate investment in scientific infrastructure has been highlighted in several recent reports sponsored by groups such as the Committee of Vice-Chancellors and Principals and the Royal Society. Estimates of the funding gap for research range from £137 million to £720 million (US\$228 million to \$1,198 million).

The Dearing committee endorses the general message of these reports, agreeing that "funds currently available to support research are barely sufficient". Indeed, one of its main recommendations is that "projects and programmes funded by the research councils meet their full indirect costs", including the costs of premises and central computing.

The report calculates that an extra £110 million is needed immediately. This would enable the research councils to increase the rate of overheads paid from 45 per cent of direct costs (excluding salaries) to at least 60 per cent, if not more where this can be justified by the institution concerned.

The main question facing the government, if it accepts these conclusions, is where this money should come from. Some are said to have told the committee that it should be taken out of the funds allocated direct to universities by the four national funding councils that allocate government grants to universities in England, Scotland, Wales and Northern Ireland respectively.

But any such move has been strongly resisted by the universities, which would lose control over the way the money was spent within their institutions if it was adopted. The universities also claim that it would undermine the viability of the so-called 'dual-support system' through which university



Dearing: seeking more government money for universities but also more selectivity in research.

research has traditionally been funded.

Reflecting these concerns — and its own determination to see the plurality inherent in the dual-support system maintained — the Dearing report lists two alternative courses of action. Its preference is that the extra money be given to the research councils as an injection of new funds. Its second option is that research councils reduce the number of research projects they support.

An additional proposal intended to address the problem of lack of funds for university research equipment is that the government should support the creation of a low-interest loan fund for such investments, with capital that could eventually approach £400 million or £500 million.

Money borrowed from this fund by "a limited number of top quality research departments" would be paid back through the overhead costs charged to research councils, government departments and other sponsors of research.

The loan fund idea has received a cautious welcome, even though there are few details on how it would work. More controversial are measures suggested by the committee to encourage some university departments to concentrate on their teaching — rather than research — strengths by penalizing unsuccessful bids for additional research funding through the funding councils' regular research assessment exercises.

The committee's proposal is that departments that agree not to bid for such funds should be given an extra £500 for each faculty member to cover 'scholarship' related activities, which could be teaching or 'private' research.

Save British Science says that it "reserves judgement" on the value of the 'scholarship' grant. But Oxburgh, who in the mid-1980s carried out the first-ever grading of university science departments, in that case in the Earth sciences, points to universities in the United States, such as Wellesley College in Massachusetts, which have successfully developed identities as primarily teaching institutions. □

Internet windfall leaves the NSF spoilt for choice

[WASHINGTON] The US National Science Foundation (NSF) has so far collected \$29 million in Internet registration fees, and the unexpected windfall will grow to \$60 million by next year, Jo Bordogna, the NSF's acting deputy director, told a Senate hearing last week.

The NSF, which funds most non-biomedical university research in the United States, does not know what to do with the money, which arises from its agreement with a private corporation, Network Solutions, to register Internet domain names.

The deal returns \$15 each year for every domain registered on the Internet to a special fund for 'Internet intellectual infrastructure'.

Bordogna told a hearing of the Senate Commerce Committee that options include returning the money to the people who paid the fees, spending it on more NSF research grants or using it as seed money for a kind of trust fund that would help to maintain the Internet. Network Solutions and NSF will decide what to do with the money early next year, Bordogna told the Senate committee.

Polly joins Dolly in the record books

[LONDON] The births of the world's first transgenic cloned lambs were announced last week by the company PPL Therapeutics and the Roslin Institute near Edinburgh, the institutions that together successfully cloned the lamb Dolly earlier this year (see *Nature* 385, 810; 1997).

One of the transgenic lambs, named Polly, carries a human gene for making a therapeutic protein, and two others are expected to carry the human gene and a marker gene. The lambs were produced using the same technique that created Dolly, except that Polly's original nucleus had been modified to carry the protein before being inserted into sheep's eggs which had their DNA removed.

Brussels wants labels on modified foods

[BRUSSELS] The European Commission in Brussels is calling for all food, seed and animal feed to carry labels signalling whether or not they contain genetically modified organisms. The commission is expected to propose legislation later this year. Observers believe the proposals are unlikely to upset the United States, as they do not call for modified foods to be

separated from unmodified ones, which US farmers bitterly oppose.

The plans have emerged in response to concern in some European countries that genetic modification may be unsafe. Austria, Italy and Luxembourg have already banned the sale of genetically modified maize.

Last month, Britain announced that plans to grow oilseed rape modified to confer tolerance to herbicides would be exposed to public scrutiny.

Algebra teaching in UK 'in need of reform'

[LONDON] The teaching of algebra in UK schools needs to be overhauled, with more emphasis placed on the process of solving algebraic equations, according to a report from a working party drawn from the Joint Mathematical Council of the UK and the Royal Society.

The report's authors say that too much emphasis is placed on teaching how to recognise and represent a pattern or relationship in algebraic terms. "The mere use of algebraic symbols does not imply algebraic activity," they say. The report recommends that algebra be introduced from the beginning of secondary school teaching. But it disagrees with calls to ban the use of pocket calculators.

IBM sets up research centre in India

[NEW DELHI] The giant computer company IBM is to establish a US\$25 million research centre in India later this year. The India Solutions Research Center will focus on infrastructure projects on weather forecasting, electronic commerce, mobile telephone systems and distance learning, in collaboration with local research centres and universities.

Alok Aggarwal, an IBM senior research manager based in the United States, will go to New Delhi as the centre's director. The initial complement of 20 scientists and engineers is expected to double by next year, rising to 100 by the next century. IBM at present has seven research centres in different countries, including the United States, Japan, Switzerland, China and Israel.

Training networks are given ECU190 million

[BRUSSELS] More than 1,100 of the leading research teams in Europe are to be brought together into 147 research training networks to be financed by the European Commission in Brussels under its Training and Mobility of Researchers Programme. The networks will concentrate on promoting the mobility and training of

young postdoctoral researchers, who will be provided with funds to study outside their home countries for several years.

The objectives of the various training networks range from developing antibiotics to the development of high-temperature superconductors. The networks will share a total budget of ECU190 million (US\$205 million), and the 147 successful bids, announced in Brussels last week, were chosen from more than 1,000 applications.

Journals link up in fight against fraud

[LONDON] The editors of a number of British medical journals, including the *British Medical Journal* (BMJ) and *The Lancet*, announced last week that they have formed an ethics committee in an attempt to combat research fraud. The committee will encourage editors to "respond more rigorously" to possible research misconduct, "rather than taking the easy option of simply rejecting questionable papers," says Richard Smith, editor of the *BMJ*.

Smith adds that the medical profession's failure to get to grips with the problem undermines its case for self-regulation. The committee is the idea of Michael Farthing, editor of the journal *Gut*, who encountered four cases of misconduct in his first year.

France seeks to open up *grandes écoles*

[PARIS] Jacques Attali, a former chief adviser to the late president François Mitterrand, has been asked by France's new minister of education, research and technology, Claude Allègre, to carry out a study of ways to establish closer links between France's universities and its elitist professional institutes, the '*grandes écoles*'.

Higher education in France has long been divided into two separate and often competing sectors. Attali, who is also a former head of the European Bank of Reconstruction and Development, has been asked to draw up suggestions for ways in which they might collaborate more closely. He will work with a panel of expert advisers and the inquiry will be broad ranging, covering not only teaching, academic awards and the mobility of students, but also the research activities carried out within both sectors.

Allègre has also set up a panel to suggest ways of tackling the problem of unemployed or under-employed doctoral graduates. At a meeting held recently in Paris, it was agreed that the panel should set up four working groups to look at different aspects of the problem. The synthesis of their conclusions will be prepared by mid-September.

US support for malaria research

Sir — I applaud the attention you have given to the Multilateral Initiative on Malaria (MIM)¹, but I wish to correct statements in your most recent News report² and leading article³ about the position of the US National Institutes of Health (NIH) on the initiative.

Although we have been prepared to consider “a common, centrally managed fund for malaria research”, at least on a small scale, the position we clearly espoused at the recent meeting in The Hague was different: a coordinated review of proposals for collaborative research involving African scientists, with funding by individual agencies according to their own priorities.

It is also inaccurate to say that we were “springing... radical proposals” on our intended partners; the several options for interactions among funding agencies had

been discussed previously in Dakar and were circulated among the agencies before the meeting in The Hague.

Furthermore, you inappropriately cast us as Jamesian upstarts, when in fact the United States supports more than half of the world's research on malaria, including substantial work in Africa⁴.

Missing from your coverage of the recent meeting, however, was the central feature of MIM to which many agencies had subscribed at a previous meeting in April 1996: the intention of developing stronger and more interactive science in Africa through efforts on malaria, in collaboration with Northern partners. I hope that, in broadening the scope of its efforts to advocacy and fund-raising, the MIM does not lose sight of that important objective.

Despite the evident resistance in The

Hague to mechanisms for multilateral support, we at the NIH are encouraged by the dedication of our partners to malaria research, eager to pursue the tasks assigned to us (proposals for a basic science agenda for malaria and for improved electronic communications in Africa), committed to further increases in fiscal support for malaria research and enthusiastic about both the original and the expanded goals of the MIM.

Harold Varmus

(Director)

National Institutes of Health,
Bethesda, Maryland 20892, USA

1. Butler, D. *Nature* **386**, 535–540 (1997).

2. *Nature* **388**, 211 (1997).

3. Butler, D. *Nature* **388**, 219 (1997).

4. *Malaria Research, An Audit of International Activity*, PRISM Report No. 7 (The Wellcome Trust, September 1996).

EMBO fellows go home

Sir — The European Molecular Biology Organisation (EMBO) is one of the organizations that award postdoctoral fellowships in molecular biology. The fellowships are funded by the 21 member states of the European Molecular Biology Conference (EMBC). Recipients have to move to a different (usually European) country for the period of the fellowship.

It is obviously important to know if these fellowships prove successful in a science career. Another question is whether the fellows remain in the country they visited or return home.

EMBO has therefore carried out a survey in which EMBO fellows from 1984–85 (approximately 200 in all) were traced by a Medline search of recent publications or followed by direct contact with the receiving institute for the fellowship or with their original home institute. The use of Medline as a primary screen for the fellows ensured that there was not a bias in favour of those who had been successful and therefore were more likely to respond to questionnaires. Given that more than 10 years had elapsed since the award of the fellowships, it was expected that the data obtained would be of relevance to their final career path. Eighty-seven per cent of the fellows were traced. Of these, 88 per cent were visibly involved in research (that is, had publications cited in Medline in the past two years) and 7 per cent were in industry.

Of particular relevance to EMBO and to the EMBC is the fact that 73 per cent of the fellows had returned and were working in their home country. This figure is particularly notable, because 29 per cent of those who received fellowships in those years carried out their research in laboratories outside Europe.

The quality of the positions of the fellows was also analysed. Eighty per cent had permanent jobs, of whom 36 per cent were in a professorial post or its equivalent in their home country. One final aspect of the survey is that it showed that the careers of the female fellows were indistinguishable from, and as successful as, those of the men.

The results of this survey reinforce EMBO's long-held belief that its fellowships contribute significantly to the scientific life of the country from which the fellow comes, and that those selected as EMBO fellows are of a very high calibre, judging by their subsequent career paths. A more detailed report providing the statistics summarized here is available on the EMBO website: <http://www.EMBO.org>

Frank Gannon

Jennifer Norman

Mare Kriis

Anne Walter

EMBO, Meyerhofstr. 1,
D-69117 Heidelberg, Germany

Lars Breimer

Clinical Research Division,
Sanofi Winthrop,
1 Onslow Street,
Guildford, Surrey, UK

Nightmare antibiotics

Sir — Harder and colleagues recently reported a new peptide antibiotic produced in human skin (*Nature* **387**, 861; 1997) — a fascinating piece of detective work. Their closing statement, however, is giving me nightmares.

They suggest that this and other endogenous human peptide antibiotics “might be ideal therapeutic agents, avoiding the problems of acquired resistance”.

Of course, another antibiotic to turn to when all else fails would, no doubt, temporarily relieve the problems of increasing resistance to the existing antibiotic armoury.

But imagine the consequences if this antibiotic were overused as others have been, and microorganisms developed resistance to it.

This is the stuff of science fiction horror — Satan bugs which have acquired resistance to an endogenous defence mechanism that has been protecting our ancestors for millions of years.

Perhaps this is one agent that should be kept where it belongs, in what is literally the last line of defence — our skin.

Chris F. Inglehearn

Department of Molecular Genetics,
Institute of Ophthalmology,
University College London,
Bath Street,
London EC1V 9EL, UK
e-mail: cinglehe@hgmpr.mrc.ac.uk

The democracy of the genes

Matt McGue

The genetic heritability of IQ remains highly contentious. A new analysis shows that genetic influences may be weaker, and prenatal environmental influences greater, than previously appreciated.

Sir Francis Galton's contributions in 1869 alone were enough to ensure that his scientific legacy would be a lasting one. In that year, he not only helped in founding this journal, but also published the first empirical investigation of the inheritance of human achievement¹, a study that is generally credited with framing the modern nature–nurture debate. On page 468 of this issue², Devlin, Daniels and Roeder report a statistical analysis of more than 200 familial IQ correlations, the most recent in a long line of empirical investigations aimed at resolving issues that were raised by Galton.

Because it is associated with a wide range of social effects, including educational and occupational success, poverty and even delinquency³, IQ has long been of interest to behavioural and social scientists. But these associations have also challenged cherished beliefs about the nature of social achievement. If social status is influenced by IQ, which is in turn substantially inherited, then social standing will, in part, be a function of one's genetic endowment. Hard work may be no guarantee of success, unless one has also received a lucky draw in the genetic lottery.

The general uneasiness over IQ reached new heights with the publication of *The Bell Curve*⁴, by Herrnstein and Murray, in 1994. The authors described declines in the population's genetic potential for high IQ (that is, dysgenic trends), because of higher fecundity among the poorly educated relative to the well educated; and they also forecast the establishment of a caste-like cognitive élite, which would be maintained through intermarriage and the (genetic) transmission of high IQs from parents to offspring. In short, they argued that Western (and in particular US) society was becoming a two-class system, where the cognitively limited masses would be ruled by a relatively small and reproductively isolated cognitive élite.

Devlin and colleagues' findings will lead to a reconsideration of these most dire conclusions from *The Bell Curve*. The likelihood of dysgenic trends occurring, as well as the ability of parents genetically to reproduce desired traits in their children, both depend on the strength of parent–offspring transmission, which in turn depends on the narrow-sense, rather than the broad-sense (total) heritability. Quantitative geneticists

distinguish additive gene effects, which are independent of genetic background and thus shared by parents and their offspring, from non-additive gene effects, which are dependent on genetic background and thus not shared by parents and offspring. Narrow-sense heritability is the proportion of a trait's variance that is attributable to additive gene effects; broad-sense heritability is the proportion attributable to both additive and non-additive gene effects.

Herrnstein and Murray⁴ based their arguments on an IQ heritability of 60%; Devlin *et al.* report a broad-sense heritability estimate of 48% and a narrow-sense heritability of only 34%. Although a narrow-sense heritability of this magnitude does not preclude dysgenic trends (albeit at a slower rate than Herrnstein and Murray projected), it is certainly too low to support the establishment of a high-IQ caste. Over the long term, unless both the narrow-sense heritability and the rate of intermarriage is very high, genes for quantitative traits such as IQ are essentially democratic; by the third or fourth generation, descendants of gifted individuals are not much more likely to be gifted than are descendants of ordinary people.

Although their paper is entitled "The heritability of IQ", Devlin and colleagues' most important finding probably concerns the nurture rather than nature of IQ. Behavioural geneticists had previously pointed out an inconsistency in the familial IQ correlations. From data on twins that have been reared together, the heritability of IQ is estimated to be about 50%; but from twins reared apart, it is estimated at about 70% (ref. 5). Because IQ correlations for twins reared together are based primarily on adolescents or children, whereas correlations for twins reared apart are based primarily on middle-

aged adults, the discrepancy had been thought to reflect age-related increases in the heritability of IQ. This hypothesis is supported by research on octogenarian twins published earlier this year⁶, as well as by studies of other familial pairings⁷.

Devlin *et al.* offer, and provide statistical support for, an alternative account of this anomaly. Specifically, they hypothesize that failure to consider the shared prenatal environment of twins results in an overestimation of IQ heritability in studies of twins reared apart, but would not bias heritability estimates in studies of twins reared together, where shared prenatal effects are cancelled out in the comparison of monozygotic and dizygotic twin similarity. When shared prenatal environments were included in the statistical model that Devlin *et al.* fitted to the IQ correlations, they accounted for 20% of IQ similarity among twins but only 5% among non-twin siblings. The former estimate is particularly remarkable given that twins, and especially monozygotic twins, can experience radically different intrauterine environments even though they share the womb at the same time⁸.

Devlin and colleagues' report supports the view that the main environmental influences on IQ occur early in life. If indeed they do, improved cognitive functioning might be an unexpected benefit of public health initiatives aimed at improving maternal nutrition and reducing prenatal exposure to toxins. Nonetheless, it is important to recognize that the evidence of Devlin *et al.*, like that supporting the influence of early intellectual stimulation on synapse formation and subsequent human intellectual performance, is indirect and as yet unreplicated.

Caution is certainly warranted. In large-scale studies where pre- or perinatal influences on IQ have been assessed directly, little evidence for any strong effect has been found⁹. Perhaps this early work lacked precise assessments of prenatal exposures, or perhaps the effect of any single prenatal factor is relatively minor and thus difficult to detect. Even though it does not establish the effect of any specific prenatal factor, Devlin and colleagues' statistical analysis implies that a relatively sizable portion of IQ variability can be attributed to the aggregate effect of the prenatal environment, and so provides a rationale for reconsidering these earlier studies.

That the IQ debate now centres on whether IQ is 50% or 70% heritable is a remarkable indication of how the nature–nurture debate has shifted over the past two decades. The anti-hereditarian position that there are no genetic influences on IQ has crumbled for want of any empirical data that would support such a radical view. Equally remarkable is the increasingly dominant view that the major environmental influences on IQ occur within the first few years of

That the debate now centres on whether IQ is 50% or 70% heritable is a remarkable indication of how the nature–nurture question has shifted

life, or in the womb, and directly affect the development of the brain. Research on the nature and nurture of IQ is converging on the view that human intellectual ability has a strong, but malleable, biological basis — a convergence that Galton would, no doubt, have found quite congenial. □

Matt McGue is in the Department of Psychology and the Institute of Human Genetics, 75 East River Road, University of Minnesota, Minneapolis, Minnesota 55455, USA.
e-mail: Mmcgue@tfs.psych.umn.edu

Palaeoclimatology

A warm future in the past

William R. Howard

To understand the impact of a possibly warmer future climate, geologists are searching the past for warmer-than-present interglacial intervals — examples of what we may expect in a world with more greenhouse gases than ours¹. A common tactic is to look at the last interglacial (around 120,000 years ago), but a period 423,000 to 362,000 years ago may fit the bill better, because the Earth's orbital geometry has been similar during the Holocene (the present interglacial period) to what it was then. This period is known to palaeoclimatologists as stage 11 or 'MIS 11', according to a numbering system for glacial advances and retreats marked in the marine oxygen isotope record (Fig. 1). At a recent symposium*, evidence emerged of extreme climatic variation and peculiar interplays between ocean temperature, thermohaline circulation, plankton ecology, sea level and reef growth during MIS 11, all of which may provide insight into the response of the natural carbon cycle to future climate change.

Changes in the Earth's orbital geometry explain the strong 41,000-year and 23,000-year rhythms in climate over the past five million years (due to changes in axial tilt and orbital precession, respectively). During the past 600,000 years, however, the dominant rhythm has been a 100,000-year cycle. Although this climate cycle is at a similar frequency to orbital eccentricity variation, the radiative impact of that orbital cycle is too small — at least in linear climate models² — to produce the large swings in the climate record.

MIS 11 is the warm peak of one such '100k' cycle, with two intriguing features. First, climate change at the 12–11 transition is nearly the most severe of the past half-million years (Fig. 1), yet it comes at a time of low orbital eccentricity, thus of low precessional amplitude. Second, interglacial conditions last longer. Subsequent interglacials (stages 9, 7, and 5) show short periods of warmth

1. Galton, F. *Hereditary Genius: An Inquiry into its Laws and Consequences* (Macmillan, London, 1869).
2. Devlin, B., Daniels, M. & Roeder, K. *Nature* **388**, 468–471 (1997).
3. Neisser, U. et al. *Am. Psychol.* **51**, 77–101 (1996).
4. Herrnstein, R. J. & Murray, C. *The Bell Curve: Intelligence and Class Structure in American Life* (Free Press, New York, 1994).
5. Plomin, R. & Loehlin, J. C. *Behav. Genet.* **19**, 331–342 (1989).
6. McClearn, G. E. et al. *Science* **276**, 1560–1563 (1997).
7. McGue, M., Bouchard, T. J. Jr, Iacono, W. G. & Lykken, D. T. in *Nature, Nurture and Psychology* (eds Plomin, R. & McClearn, G. E.) 59–76 (Am. Psychol. Assoc., Washington DC, 1993).
8. Price, B. *Am. J. Hum. Genet.* **2**, 293–352 (1950).
9. Broman, S. H., Nichols, P. L. & Kennedy, W. A. *Preschool IQ: Prenatal and Early Developmental Correlates* (Erlbaum, Hillsdale, NJ, 1975).

followed by early returns to glacial conditions (Fig. 1).

Many other aspects of the climate were strange. In deep-sea records of MIS 11, sub-polar waters in both hemispheres were at their warmest (L. Burckle, Lamont-Doherty Earth Obs.), following one of the coldest glacial stages of the past 500,000 years (Fig. 1c, d). Deep-ocean carbon isotope data show a maximum in the production of North Atlantic Deep Water (D. Hodell, Univ. Florida). Tracers of downward particle flux in the equatorial Pacific show that biological productivity was at a minimum in MIS 11 (R. Murray, Boston Univ.). Calcium car-

bonate production peaked in subpolar and subtropical oceans (R. Schneider, Univ. Bremen), reflecting a temperature-related shift in plankton ecology from diatoms to calcareous plankton.

In shallow-water environments, the evolution of several major reef systems (such as the Great Barrier Reef) accompanied increased reef carbonate production (A. Droxler, Rice Univ.; P. Davies, Sydney Univ.). The production of carbonate in continental shelves and in mid-latitude open-ocean environments may partly explain the extreme dissolution of carbonate sediments during MIS 11 throughout ocean basins such as the Indian and Pacific³. (In the oceans' carbonate cycle, biological formation of hard body parts from carbonate outstrips river input of the component ions, and dissolution from the seabed makes up the difference. In other words, to produce more carbonate in one place you must dissolve more elsewhere.)

Perhaps barrier-reef tracts were established during stage 11 because tropical continental shelves were flooded (A. Droxler; C. Kievman, Kean Coll.). How high did the seas rise? Beach deposits in Alaska, Bermuda and the Bahamas (P. Hearty, Bahamas) and uplifted reef terraces in Indonesia⁴ seem to imply that the ocean stood as much as twenty metres above its present level. But this is uncertain: estimating sea level for MIS 11 is complicated by the fact that dating beaches

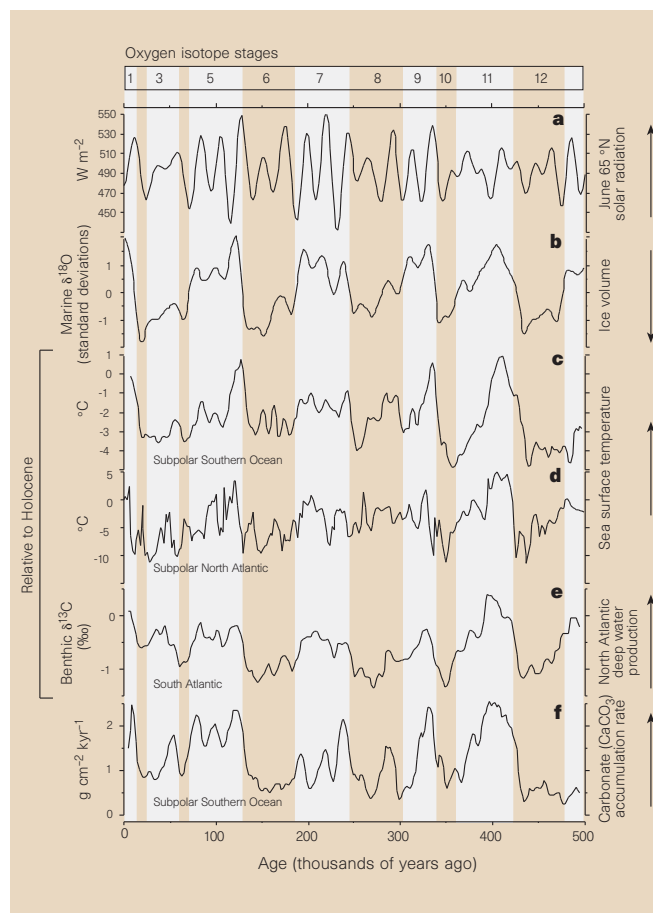


Figure 1 Climate and chemistry of the past 500,000 years. All the climatic variables b to f (refs 2, 8–11, respectively) are related to solar forcing, a (ref. 12), determined by changes in the Earth's orbit (c, d and e relative to average over the Holocene — the past 10,000 years or so). The long, hot interglacial of stage 11 occurred under similar orbital conditions to those we have now — so is it a model for our own time?

*Carbonate Marine System During Oxygen Isotope Stage 11, American Geophysical Union Spring Meeting, Baltimore, Maryland, USA, 27–30 May 1997.

and reefs of that age is difficult. Deep-sea $\delta^{18}\text{O}$ records (including some new data presented at the symposium) show isotopic depletions during MIS 11 that are consistent with a sea-level high, but in the isotope records the effects of warmer temperatures cannot be easily untangled from the effects of decreased ice volume.

To produce such a high sea level would require the collapse of at least one major ice sheet (candidates include the West Antarctic and Greenland Ice Sheets). The stability of these ice sheets in the face of high-latitude warming is one of the main questions in climate-change research.

These extremes in the ocean's thermal and geochemical state all lower its ability to hold CO_2 , leaving more in the atmosphere. First, CO_2 is less soluble in a warmer ocean. Second, as sea level and surface temperature rise, a greater proportion of biological production is in the form of calcium carbonate. This increased production and burial of calcium carbonate lowers the ocean's alkalinity, again reducing its ability to hold on to CO_2 . Third, decreased biological production of soft tissue reduces the export of carbon from the surface (in contact with the atmosphere) to the deep ocean⁵. These changes in the ocean's carbonate-carbon cycle during MIS 11 point to what might be an important set of interactions between temperature, ecology and CO_2 . Model simulations of oceanic CO_2 uptake have recently begun to take such feedbacks into account⁶.

Was the stage-11 warming driven by increased atmospheric CO_2 concentrations? The Vostok ice core now provides a palaeoclimate record extending to about 425,000 years ago⁷. Preliminary measurements suggest that atmospheric CO_2 levels during stage 11 are comparable to those of later interglacials, which would mean that the link between warming mechanisms and the carbon cycle may not be so simple. However, ice disturbance at the stage-11 level may have altered the CO_2 record (D. Raynaud, LGGE, Grenoble).

The symposium addressed only some of the evidence from and implications of the stage-11 warm interval. Was the climate stable, or punctuated by high-frequency 'flickers'? What was the global distribution of sea-surface temperature? Stage 11 has not been studied in detail, in part because few climate records have reached it. But new cores, extracted by Ocean Drilling Program technology and long gravity-driven piston cores in high-accumulation-rate drift deposits, are beginning to provide more evidence.

Stage 11 may provide some important perspectives on future climate change. Are we adding greenhouse gases to a climate that, left to its own devices, should be on the way into another glaciation, as in the short-lived stage 5; or one that would stay warm, or get

warmer, for another 40,000 years, as in stage 11? Are we setting in motion a chain of biogeochemical feedbacks that will add an oceanic contribution to the anthropogenic greenhouse-gas enrichment? □

William R. Howard is at the Cooperative Research Centre for the Antarctic and Southern Ocean Environment, University of Tasmania, GPO Box 252-80, Hobart, Tasmania 7001, Australia.

1. Burckle, L. H. *Quat. Sci. Rev.* **12**, 825–831 (1993).
2. Imbrie, J. *et al.* *Paleoceanography* **8**, 699–735 (1993).
3. Farrell, J. & Prell, W. L. *Paleoceanography* **4**, 447–466 (1989).

4. Pirrazoli, P. A. *Mar. Geol.* **109**, 221–236 (1993).
5. Broecker, W. S. & Peng, T.-H. in *The Global Carbon Cycle* (ed. Heimann, M.) 95–115 (Springer, New York, 1993).
6. Sarmiento, J. & Le Quéré, C. *Science* **274**, 1346–1350 (1996).
7. Petit, J. R. *et al.* *Nature* **387**, 359 (1997).
8. Howard, W. R. & Prell, W. L. *Paleoceanography* **7**, 79–118 (1992).
9. Ruddiman, W. F., Raymo, M. E., Martinson, D. G., Clement, B. M. & Backman, J. *Paleoceanography* **4**, 353–412 (1989).
10. Oppo, D. W., Fairbanks, R. G., Gordon, A. L. & Shackleton, N. J. *Paleoceanography* **5**, 43–54 (1990).
11. Howard, W. R. & Prell, W. L. *Paleoceanography* **9**, 453–482 (1994).
12. Berger, A. L. *Quat. Res.* **9**, 139–167 (1978).

Protein engineering

Reading, writing and redesigning

Michael Groß and Kevin W. Plaxco

Nature doesn't have a protein-folding problem — but we do. Although we know that the line-up of the amino-acid building-blocks in a protein encodes the architecture of its correctly folded (native) form, we still cannot properly read or write this code. That is, we find it difficult to predict structures from sequences, or to design sequences that fold into a desired structure. In 1994, this dilemma prompted George Rose and Trevor Creamer¹ to offer a cash prize for the solution to a specific protein-design problem. And in this month's issue of *Nature Structural Biology*, Lynne Regan and co-workers report² that they have redesigned a protein into an entirely different fold by exchanging only 50% of its amino acids. This remarkable achievement has earned them the prize, less than four years after it was announced.

Engineering prizes have, arguably, driven technological advances in the past — from the first human-powered flight across the English Channel (Kremer prize), to the first steps in nanotechnology (Feynman prize). The protein-design equivalent of the first cross-Channel flight is the 'Paracelsus chal-

lenge', named after the sixteenth-century Swiss physician, philosopher and alchemist. In an editorial in *Proteins*, Rose and Creamer proposed the challenge, pledging \$1,000 for the first person to show a pair of proteins which have distinctly different topologies, but which retain no less than 50% sequence identity. Such pairs are not found in nature — databases show that pairs of proteins sharing 25% of their amino-acid residues usually have very similar structures³.

An identity score of 50% between two natural protein sequences would convince any biochemist that the proteins must be nearly identical in structure. On the other hand, because the structural identity of a protein is determined by roughly a quarter of its building blocks, the task — which is a bit like turning this article into a report about a tennis match by exchanging only half of the words — is a realistic one. In the language of proteins, most of the words seem to be space fillers that do not convey a specific meaning, so the challenge is to find the important words and rewrite them. Regan and co-workers have met the challenge with style. Rather than simply converting a given pro-

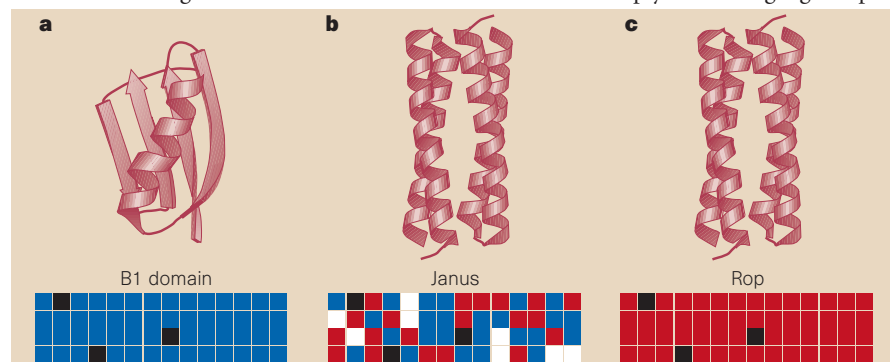


Figure 1 The two faces of protein folding. Dalal *et al.*² have risen to meet the 'Paracelsus challenge', by designing a pair of proteins that have completely different folds yet have no less than 50% sequence identity. a, The starting material was the B1 domain of the streptococcal IgG-binding protein G, which is mainly β -sheet. b, By modifying just 28 of the 56 amino acids in this region, they created a new protein which they called Janus. Janus has 50% sequence identity with the B1 domain, yet its mainly helical structure resembles that of the RNA-binding protein Rop. c, Rop, a homodimeric four-helix bundle protein. Blue boxes, residues from B1; red boxes, residues from Rop; black boxes, residues that are identical in B1 and Rop; white boxes, residues from neither B1 nor Rop.

tein into whatever different structure was the most easily attainable, they specifically set out to convert the B1 domain of protein G, which consists of a four-stranded β -sheet packed against a single helix⁴, into a new protein with the folded structure of another protein, Rop (repressor of primer). Rop is an RNA-binding protein that folds as two helices associating to form a four-helix bundle⁵. Their strategy was to identify the key residues of the B1 domain that encode its mostly β -sheet structure, and to replace those residues with the corresponding, structurally critical amino acids of Rop.

What are the essential features of a protein that we have to specify to define its fold? There are four important goals, all of which aspiring protein designers must achieve⁶. The first is to initiate and terminate elements of secondary structure (the helices and β -sheets that are shown as ribbons in structure cartoons; Fig. 1). The second is to create a hydrophobic core — a congregation of water-avoiding amino-acid side chains whose reluctance to be exposed to the solvent will stabilize the correctly folded structure. The third goal is to supply sequences that encode loops and hairpin turns in the protein backbone, to connect sequential, secondary-structural elements. Finally, specific tertiary links need to be created; that is, energetically favourable interactions between amino acids that are distant in the sequence.

Following this logic, to swap the protein identities Regan and co-workers first had to replace residues that had an empirically observed tendency to form sheets, with other residues that prefer to be in a helix. Then they constructed the appropriate core structure based on their previous characterization of simplified core variants of Rop^{7,8}. They also pinched from the Rop sequence a salt bridge (a pair of oppositely charged amino acids) and a single tyrosine residue as a spectroscopic probe of the folded state.

These modifications account for 28 amino acids — exactly 50% of the original B1 sequence. As well as the 56 residues that form the helices (and were the design target in this study), Rop has an unstructured 'tail' of seven residues. This was found to be essential for the good water solubility of Rop, although not for its structure, and it was also included in the design.

The new protein was fittingly called Janus (Fig. 1), as it looks like the B1 domain on the sequence level but was expected to show the face of Rop on the structural level. Regan's group cloned Janus into bacteria so they had enough material for structural investigations and could address the crucial question — does Janus really look both ways? As they had predicted, Janus showed the circular dichroism and nuclear magnetic resonance signatures, isotope-exchange properties and highly cooperative unfolding that would be expected of a folded protein with the structure of Rop.

Based on these results, it was agreed that Janus had met the criteria of the challenge, and Regan was awarded the Paracelsus prize in a small ceremony last month.

Although this achievement is an impressive demonstration of just how far our knowledge of the determinants of protein structure has come, a few questions remain. For example, can the limit of sequence identity be pushed higher than 50%, and can the transformation be done in the opposite direction? Regan's group is now addressing the limits of the sequence identity by reverting the changes made from B1 to Janus, one by one. But it might be more difficult to do the transformation in the opposite direction, as the formation of β -sheet structures depends on interactions between sites that may be far apart in the sequence. Still, the demonstration that

protein designers now know which half of the text they have to rewrite to change a protein fold means that, one day, we will be able to read and write the code of protein folding. □

Michael Groß is at the Oxford Centre for Molecular Sciences, Oxford OX1 3QT, UK. Kevin W. Plaxco is in the Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA.

1. Rose, G. D. & Creamer, T. P. *Proteins Struct. Funct. Genet.* **19**, 1–3 (1994).
2. Dalal, S., Balasubramanian, S. & Regan, L. *Nature Struct. Biol.* **4**, 548–552 (1997).
3. Chothia, C. & Lesk, A. M. *EMBO J.* **5**, 823–826 (1986).
4. Gronenborn, A. M. *et al. Science* **253**, 657–661 (1991).
5. Banner, D. W., Kokkinidis, M. & Tsernoglou, D. *J. Mol. Biol.* **196**, 657–675 (1987).
6. Cordes, M. H. J., Davidson, A. R. & Sauer, R. T. *Curr. Opin. Struct. Biol.* **6**, 3–10 (1996).
7. Munson, M., O'Brien, R., Sturtevant, J. M. & Regan, L. *Prot. Sci.* **3**, 2015–2022 (1994).
8. Munson, M. *et al. Prot. Sci.* **5**, 1584–1593 (1996).

Sociology of science

Irregular scientific conjugations

The perception of behaviour varies with the distance of the observer from the subject. This variation generally extends along a continuum but, in the special case of the behaviour of scientists, three primary clusters of distance seem particularly important. These can be defined as identity (I), unrelatedness (U) and separated/heterologous (S/he). Results in ten areas of behavioural perception within science can thus be summarized as irregularly conjugating verbs.

I construct precise and logical sentences

You write in a standard scientific style
She produces incomprehensible papers riddled with jargon

I am frequently asked to appear in TV science documentaries

You are a typical eccentric media-friendly boffin
He is a self-promoting pseudoscientific media whore

I am a polymath

You are a jack of all trades
She has views on subjects she knows nothing about

I publish every finding for the wider scientific community

You have a refined strategy for building an impressive publications list
They are salami slicing

I am a confident speaker in panel discussions

You are charmed by the sound of your own voice
He shouts down the views of other participants

I have a reputation that enables me to attract bright young scientific talent

You have a reputation maintained by a highly able team of collaborators
He has a reputation for exploiting poorly paid postdocs

I ensure that a company's scientific approach is up to date

You reassure investors by your presence on the Scientific Advisory Board
He receives a fat retainer simply because he has a Nobel prize

I undertake peer-review carefully and thoroughly

You delay returning your comments on the paper
She conducts additional experiments for her own paper while the manuscript languishes in her pending tray

I do not publish preliminary data

You conduct follow-up experiments in order to maintain a scientific lead over competitors
He hoards data until his industrial collaborators have filed patents

I actively solicit and review high-quality papers

You provide useful comments on the occasional manuscript
They are only on the editorial board to preserve the façade of geographical balance

John Hodgson

John Hodgson is a senior editor of *Nature Biotechnology*.

Quantum mechanics

Whistles from superfluid helium

Peter McClintock

On page 449 of this issue¹, Pereverzev *et al.* report mechanical oscillations spontaneously generated by two weakly coupled volumes of superfluid ³He—providing a direct demonstration of its large-scale quantum behaviour.

At very low temperatures all materials tend to fall into highly ordered states. Most commonly, this order takes the form of perfect crystals, but there are exceptions. Most strikingly, the two isotopes of helium (³He and ⁴He) never solidify at all under their vapour pressures. Rather, they stay liquid to the lowest temperatures attainable, and are assumed to remain so right down to absolute zero.

This is baffling at first sight, because liquids are usually disordered, whereas the third law of thermodynamics says that, as the temperature tends to zero, the entropy of any system (that is, its disorder) should decrease towards a constant value that can be taken as zero. But ordering does also occur in the liquid heliums, albeit in momentum space rather than real (cartesian) space: they both undergo a form of Bose–Einstein condensation, in which the ⁴He atoms (or pairs of ³He atoms) all fall into the same quantum state, described by a macroscopic wavefunction that fills the entire volume of liquid.

By analogy with the microscopic worlds of atoms or electrons, conditions at any point in the liquid are determined by a wavefunction. Liquid motion is specified by the relative phases of the wavefunction at different points: if its phase is the same everywhere, the liquid is stationary; if in a particular direction the phase progressively advances, then the liquid is flowing in that direction. All of this makes good sense in relation to a single bath of liquid helium, or to the electron gas (also described by a macroscopic wavefunction) in a single piece of superconducting metal. But what happens when there are two separate baths, or two separate pieces of superconductor, and they are brought together?

Based on ideas introduced by Josephson², Anderson³ and Feynman⁴, one can expect a variety of counter-intuitive and very striking phenomena to occur when two separately phase-coherent systems of this kind are weakly coupled together. One of these, the a.c. Josephson effect, consists of an oscillatory flow of particles to and fro through the weak link coupling the two systems. It occurs when the chemical potential μ (an average energy per particle) differs on each side. In the case of weakly coupled superconductors, the effect can be detected through the electromagnetic radiation that is created by the

oscillatory current of negatively charged electrons. It is inherently much harder to demonstrate in liquid helium, partly because the oscillatory current is neutral and therefore creates no electromagnetic radiation, and partly because the weak link between the baths has to be made extraordinarily small.

Pereverzev *et al.*¹ have created their weak link by using modern silicon microfabrication techniques, enabling them to construct 100-nm-diameter apertures in a 50-nm-thick sliwer of silicon nitride. The small diameter and length are crucial to success: unless they are both comparable with, or smaller than, the healing length ξ of the superfluid (the minimum length over which the wavefunction can vary significantly), the aperture cannot be expected to function as a weak link. Given that $\xi \sim 0.1$ nm for ⁴He, and ~ 50 nm for ³He, it is clear that the state-of-the-art Berkeley technology is just good enough to see Josephson effects in superfluid ³He. Because oscillatory flow in such a tiny hole would be so slight as to be undetectable, they used 4,225 identical holes, hoping that this would amplify the effect.

The silicon chip with its tiny holes was placed in the pill-box-shaped cell sketched in Fig. 1b on page 449¹. A 'soft membrane' divides the cell into inner and outer chambers, and can be moved by changing the voltage on an electrode, thereby applying an electrostatic force. Its position can be detected by means of a d.c. SQUID (a superconducting quantum interference device, used here as a position transducer of extreme sensitivity). Following a change in the position of the soft membrane, two things happen. First, the tiny change of pressure of the helium in the inner reservoir, as compared with that in the outer one, will change their relative values of μ , leading to oscillatory flow through the apertures. According to the Josephson relation, its frequency should be 183.7 kHz Pa⁻¹ for ³He. Second, superimposed on the oscillation, there will be a steady flow through the apertures that reduces the pressure difference, leading to a continuous decrease towards zero in the frequency of the oscillatory flow.

The experimenters found that the amplified output of the displacement transducer, passed to audio headphones, could be heard as a 'whistle' of steadily decreasing pitch—precisely what was expected. Remarkably, it was the discrimination and flexibility of the human ear that led to the initial discovery. Only later were the authors able to make quantitative measurements of the frequency as a function of pressure difference between



100 YEARS AGO

One is apt to forget how strong is the instinct to shout and sing, laugh and cry. It is especially noticeable in the savage and in the child. If these instincts are unduly repressed in the child, he is sure to suffer. Crying should certainly be restrained within limits, but there can be no doubt that it is primarily physiological, not only favouring the proper expansion of the lungs and accelerating the circulation, but deadening the effects of pain and relieving nervous tension (especially in woman). Rosbach thinks it not improbable that many evils which manifest themselves in later life, such as chlorosis, contracted chest and the phthisical habit, "may take their origin in the practice of mothers to stop their infants from screaming by soothing them to sleep in their arms or by stupefying rocking in the cradle." ... It is well known that children show a strong instinct to chatter and sing the first thing in the morning, and it should be allowed full vent as far as is practical. The shouting which young people indulge in during their play is quite remarkable and is manifestly physiological. The same tendency to shout is observed in young adults, especially among the poorer classes in holiday time. Though from the physiological point of view justifiable, and even beneficial, the noises they make are certainly not always pleasing, especially to the sensitive nerves of the cultured, amongst whom this instinct is consequently suppressed...

From *Nature* 29 July 1897.

50 YEARS AGO

In the course of cytological investigations for the establishment of the chromosome atlas of the endemic flora of Portugal, we found in *Luzula purpurea* Link – a rare annual species only known from Portugal, Madeira and the Canary Islands – the diploid number $2n=6$. Besides their low number, the chromosomes of this species seem to be provided with two centromeres, one placed at each end of the chromosome. No median, sub-median or sub-terminal constriction could be distinguished either in mitotic or meiotic metaphases. ... This appears to be the first time such chromosome behaviour has been found in plants, and we think it may be related to what Toledo Piza reported for the Brazilian scorpion *Tityus bayensis* Perty and for some Hemiptera. From *Nature* 2 August 1947.

the reservoirs. Gratifyingly, these fall on a straight line whose gradient agrees with the theoretical prediction based on the Josephson relations.

The way is now open for seeking other mechanical macroscopic quantum phenomena and applications of the Josephson relations in superfluid ^3He — the mechanical analogues of electrical effects in superconductors — and, perhaps, for developing

a pressure transducer of extraordinary sensitivity. □

Peter McClintock is in the Department of Physics, Lancaster University, Lancaster LA1 4YB, UK.

1. Pereverzev, S. V., Loshak, A., Backhaus, S., Davis, J. C. & Packard, R. E. *Nature* **388**, 449–451 (1997).
2. Josephson, B. D. *Phys. Lett.* **1**, 251–253 (1962).
3. Anderson, P. W. in *Lectures on the Many-Body Problem* (ed. Caianello, E. R.) (Academic, New York, 1964).
4. Feynman, R. P. *The Feynman Lectures on Physics* Sec. 21.9 (Addison-Wesley, London, 1965).

Cellular secretion

Now you see it, now you don't

W. J. Betz and J. K. Angelson

'Seeing is believing' is a reasonable guide for everyday experience. But when looking at the events that occur in living cells — events that must be observed through a light microscope — 'seeing is belying' may be a better aphorism. This is partly because photons (the fundamental unit of light) peek around edges, to diffract and blur an otherwise crisp image. Now, on pages 474 and 478 of this issue, Steyer *et al.*¹ and Ryan *et al.*² extend our ability to image individual secretory granules and vesicles in living cells, and they reveal new information about exocytosis and granule transport.

Because of optical diffraction, the smallest membrane-bound cellular organelles (such as synaptic vesicles) would appear larger than life, if they were visible at all. It's a bit like looking at a tennis ball through fogged glasses, and seeing instead a fuzzy football. In fact, it's much worse — imagine looking into a bucket of tennis balls, each one of which is transparent. This describes the biological challenge that has been elegantly addressed, and met, in these two very different papers. Ryan *et al.*² stained one (or a few) synaptic vesicles and illuminated the entire cluster (Fig. 1), whereas Steyer *et al.*¹ stained all secretory granules in chromaffin cells, but illuminated only a few (Fig. 2).

Until the early 1980s, about the only method that was suitable for studying secretion in individual living cells was the intracellular recording of synaptic signals. Then, three new techniques were developed which allowed secretion to be measured in isolated cells — a postsynaptic detector cell was no longer required. Patch-clamp capacitance³ detects exocytic and endocytic events by electrically measuring the surface area of a cell; amperometry⁴ detects secretions using a carbon electrode positioned near the cell; and optical microscopy is useful for tracking organelles labelled with fluorescent dyes, particularly dyes that are taken up in an activity-dependent fashion^{5–7}.

Staining a single synaptic vesicle is no trick, but imaging and identifying it is. A single electrical stimulus applied to a nerve terminal usually releases about one quantum

of transmitter⁸. This probably reflects exocytosis of one vesicle (from the many in the terminal), which is then endocytosed and recycled into the vesicle cluster. If the fluorescent dye FM1-43 is present in the bathing fluid during the endocytic step, the recycled vesicle becomes stained with a few hundred dye molecules⁹.

Imaging the stained vesicle is then the challenge. Ryan *et al.*² took pains to reduce background fluorescence, used a bright light for imaging (to reduce noise), and then took three pictures of hippocampal cultures in which synaptic contacts were abundant (Fig.

1). Image number one (see Fig. 1b on page 479) was obtained after giving one, or a few, electrical shocks in the presence of FM1-43, then washing thoroughly to get rid of any dye not taken up by endocytosis. This image gave a lot of really dim spots. Image number two (Fig. 1d on page 479) was acquired after exhaustive stimulation designed to release and wash away any FM1-43 that was taken up initially by recycling vesicles. Naturally, this image was even dimmer than the first. Image number three (Fig. 1g on page 479) was obtained after re-staining with FM1-43 using many electrical shocks to label all of the synaptic vesicles, so it was easy to identify the position of synaptic sites.

For the analysis, Ryan *et al.* worked backwards (see also ref. 10). They started with image number three to identify synaptic sites, then they analysed the data from images one and two at those positions. Their results indicated 'quantal' staining of synaptic sites — fluorescent spots stained after one or a few shocks showed mean brightnesses of about 0, 60, 120 or 180 photoelectrons at different synapses, probably reflecting the uptake of FM1-43 by 0, 1, 2 or 3 recycling vesicles, respectively. (Individual vesicles in the bright clusters could not be identified spatially because diffraction blurred them together.) Fluorescence at the fully stained

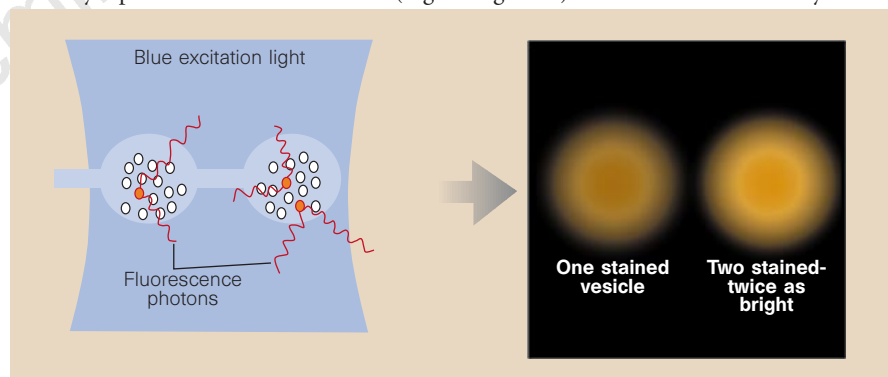


Figure 1 Stain a few vesicles, illuminate all. Ryan *et al.*² stained one, or a few, synaptic vesicles with the fluorescent dye FM1-43, by giving 1–3 electrical shocks to hippocampal neurons in culture. The brightness of different synaptic sites varied in a stepwise fashion, each step representing one stained vesicle.

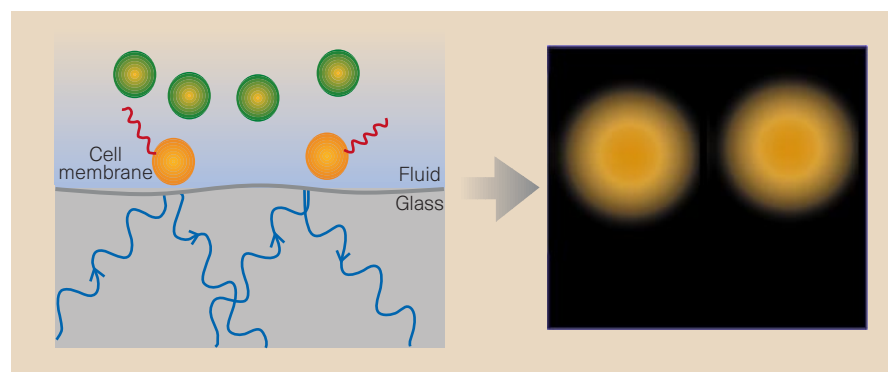


Figure 2 Stain all granules, illuminate only a few. Steyer *et al.*¹ stained all secretory granules in a chromaffin cell with acridine orange, and then illuminated only those granules that lay close to the coverslip by using the evanescent wave from a reflected laser beam.

terminals was about 127 times greater than the unit brightness, in good agreement with electron-microscopic observations, which revealed 100–200 vesicles per synapse.

Adrenal chromaffin granules are about 250 nm in diameter — considerably larger than synaptic vesicles (around 50 nm in diameter). Although this means that chromaffin granules are not as bloated by diffraction, they do not seem to recycle as faithfully as synaptic vesicles, making it impossible to stain one, or a few, of them with FM1-43. But almost all of the granules can be stained using a fluorescent dye; for example, acridine orange, which diffuses across all membranes and becomes trapped in acidic compartments such as secretory granules. However, the result is not very useful for ordinary microscopy: when all 25,000 granules in a chromaffin cell are stained, the cell fluoresces brightly and hazily, making it virtually impossible to track single granules. Even confocal microscopy does not reduce out-of-focus fluorescence enough, and the laser illumination in a confocal microscope can be quite toxic. Enter selective illumination, through evanescent-wave microscopy.

Steyer *et al.*¹ selectively excited dye that was contained in granules which lay within a few hundred nanometres — about the diameter of a single secretory granule — of the glass coverslip to which the host chromaffin cell was attached (Fig. 2). They excited only selected dye molecules using evanescent-wave microscopy¹¹. A laser is projected from below at an angle that causes the laser beam to be reflected at the glass–fluid interface, back into the glass. However, the reflected beam probes slightly into the overlying aqueous medium, and this ephemeral ‘testing of the waters’ delivers the excitatory stimulus to nearby fluorophores, with little cellular photodamage. Lest this begin to sound dangerously close to the obverse maxim ‘believing is seeing’, Steyer *et al.* combined electron microscopy, interference reflection microscopy and amperometry, to confirm that their observations were correct.

And the observations of single secretory granules were truly beautiful. Granules cruised from above into view at a rate of 114 nm s⁻¹, bumped around the surface awhile, docked (sometimes reversibly), and exocytosed with a puff as the dye diffused out of the fusion pore. Replacement with ‘reserve’ granules was slow, requiring several minutes. These findings are interesting and exciting in their own right. Moreover, the technical refinements will be especially welcome to molecular biologists, who have identified a daunting inventory of proteins that participate in these biological processes (especially docking and exocytosis), but the exact functions and interactions of which are only partly understood¹². Refined optical technologies will become useful tools in this enterprise, and cell-free assay systems — in

which the roles of individual proteins can be examined with exquisitely sensitive optical techniques such as those described here — could now be a real possibility. □

W. J. Betz and J. K. Angleson are in the Department of Physiology, University of Colorado Medical School, Denver, Colorado 80262, USA.

1. Steyer, J., Horstmann, H. & Almers, W. *Nature* **388**, 474–478 (1997).
2. Ryan, T. A., Reuter, H. & Smith, S. J. *Nature* **388**, 478–482 (1997).
3. Neher, E. & Marty, A. *Proc. Natl Acad. Sci. USA* **79**, 6712–6716 (1982).

4. Wightman, R. M. *et al. Proc. Natl Acad. Sci. USA* **88**, 10754–10758 (1991).
5. Lichtman, J. W., Wilkinson, R. S. & Rich, M. M. *Nature* **314**, 357–359 (1985).
6. Betz, W. J. & Bewick, G. S. *Science* **255**, 200–203 (1992).
7. Betz, W. J., Mao, F. & Smith, C. B. *Curr. Opin. Neurobiol.* **6**, 365–371 (1996).
8. Stevens, C. F. & Wang, Y. *Neuron* **14**, 795–802 (1995).
9. Henkel, A., Lübke, J. & Betz, W. J. *Proc. Natl Acad. Sci. USA* **93**, 1918–1923 (1996).
10. Murthy, V. N., Sejnowski, T. J. & Stevens, C. F. *Neuron* **18**, 599–612 (1997).
11. Stout, A. L. & Axelrod, D. *Appl. Opt.* **28**, 5237–5242 (1989).
12. Zucker, R. S. *Neuron* **17**, 1049–1055 (1996).

Schizophrenia

The price of language?

John Maddox

Linguistics and brain anatomy, recently become bedfellows, are destined to have even more to say to each other than they did at a workshop held in Oxford at the end of last month*. The hidden agenda, which nobody attempted to conceal, was organizer Tim Crow's hypothesis that schizophrenia stems from a failure to develop the normal asymmetry of the brain represented by the presence of a language processing centre (Broca's region) in the temporal region on the left (in right-handed people).

One thing was plain at Oxford: Noam Chomsky (who was not present) has firmly established his point that grammar, and in particular syntax, is innate. Interested linguistics people (what is the collective noun?) are busily speculating on how the language function could have evolved, presumably no earlier than *Homo erectus* (on brain-size grounds) and possibly even later than mitochondrial Eve, say 150,000 years ago. What clues are there in the modern use of language?

The general opinion at the meeting was that carrying a lexicon in the head is a secondary problem. But the analysis of speech requires a sense of syntax, which requires computation of a special kind. Derek Bickerton (Univ. Hawaii) insists that this faculty must have come into being all at once. The meaningful units of speech (or text) are clauses, and their analysis is irreducible — you cannot even tell what is the subject of the verb (or even what word is the verb) until you have heard the whole clause.

How could the language function have emerged in this way, catastrophically as it were? Most simply, runs one answer, by adapting pre-existing neuronal circuitry for the analysis of syntax. Bickerton's bet is that the hijacked circuitry is still used by modern monkeys for calculating their social relationships with their fellows. Whatever the truth, Robert C. Berwick (one of Chomsky's

colleagues at the Massachusetts Institute of Technology) argued (by demonstration) that there are computational schemes by which clauses can be parsed recursively, as if by adding successive words in a clause to a provisional analysis of the syntax. No doubt there will soon be computer programs to test Bickerton's assertion.

Handedness (left or right) is linked with the asymmetry of the speech centre, but not straightforwardly. It is not simply that the data are confounded by the attempts of fond parents and less-fond schoolteachers to make naturally left-handed children use their right hands instead, that individuals' preferences are not black and white, or that data are often gathered by means of questionnaires, but that even copper-bottomed data imply a high degree of randomness in the development of individual handedness.

Of the three theories in the field, that due to Marian Annett (Univ. Leicester) seems to have the strongest wind behind it. Hand preference is not a discrete variable, but rather is normally distributed about a mean, which is itself shifted towards the right (by about one standard deviation of the gaussian curve) as a presumed result of the asymmetry of the speech centre — which is supposed to be determined by a developmental gene. That can explain why a quarter of monozygotic twin pairs differ in handedness, and why the proportion is not much less in dizygotic twins.

One of the surprises of the workshop was the use of archaeological data in the study of handedness. James Steele (Univ. Southampton) has collected data from human skeletons to show the prevalence of right-handedness in ancient populations — the long bones of the most-used arm are lengthened. By this test, the anatomy of the occupants of a mediaeval cemetery in Yorkshire agrees almost exactly with Annett's more modern data. Steele also explained that six out of six Neanderthal skeletons sampled seem to have been right-handed, as does the Leakey–Walker *H. erectus* youth now known after the

*The Evolution of Language and the Speciation of *Homo sapiens*, The POWIC (Prince of Wales International Centre) Workshop, St Anne's College, Oxford, UK, 30 June 1997.

site of its discovery at Nariokotome in Kenya. A further source of proxy data is the working of ancient flint-knapping sites: the products of left-handed knappers are recognizable as such, as is the pattern of wear on tools used by left-handed early people.

This is not as irrelevant to the workshop's hidden agenda as it may seem. Thus, Crow described his analysis of the 11,600 children tested for hand skills and various cognitive abilities at age 11 in the British National Child Development Study; the striking conclusion is that there is a marked (and statistically significant) deficit of performance on four tests among those with equal hand skills, called 'the point of hemispheric indecision'. Ambiguity about handedness seems to be an innate disadvantage (at least at the age of 11). The same group of children is also a commoner source of schizophrenic breakdown later in life than would be expected.

So what is the gene? Its function, says Crow, must be to induce the development of the cerebral asymmetry that usually puts Broca's area on the left; a mutation would then account for its being sometimes on the

right. Post-mortem studies, and now magnetic resonance imaging, suggest that people with schizophrenia tend to lack pronounced asymmetry, making the high prevalence of the disease (apparently unreduced by natural selection despite the reduced fertility of people with schizophrenia) part of the price that *Homo sapiens* has to pay for the faculty of language.

And where is the gene? Crow's suspicion that it is X-linked is supported by the work of D. H. Skuse (Univ. College, London), who told the workshop of his work with patients suffering from Turner's syndrome (single-X females; see *Nature* **387**, 705–708; 1997) which suggests that a maternally imprinted X-chromosome gene is responsible for their measurable cognitive impairment. From this and other studies, Crow concludes that there is a gene responsible for developmental asymmetry near the centromere of the X chromosome, and also on the Y. One suspects he will not be the only one looking for it in the years ahead. But only zealots believe that such a tangled tale can have such a simple resolution. □

John Maddox is Emeritus Editor of *Nature*.

Photochemistry

Trees to trap photons

Shaul Mukamel

Photochemical reactions are usually triggered by a high-energy ultraviolet or visible photon that delivers the necessary energy. In some cases such reactions can instead be induced by the absorption of many low-energy infrared photons using intense infrared lasers. But on page 454 of this issue¹, Jiang and Aida report a remarkable observation that could lead to a new strategy for multiphoton chemistry. They attach a photoisomerizable molecule to the centre of a spherical cavity formed by a highly branched molecule, or dendrimer. Low-power infrared light then triggers the *cis*- to *trans*- isomerization (twisting the molecule around the nitrogen–nitrogen double bond).

In chemical reactions, vibrational energy affects rates and branching ratios among the various products². Unimolecular reactions such as isomerization are usually slow compared with vibrational energy flow within a molecule — vibrational energy is typically redistributed in less than a picosecond. Such reactions can therefore be described using statistical theories which assume that vibrational energy is distributed thermally among the various modes.

To induce a reaction, enough energy is needed to break a molecular bond. A single visible or ultraviolet photon is enough, but infrared photons carry a relatively small amount of energy, so in order to undergo the

same reaction a molecule needs to absorb a large number of infrared photons.

The dream of multiphoton laser chemistry has been to pump many quanta of light into a single mode (or bond) using high-power lasers, with the hope that the particular bond will break before the energy can dissipate to

other degrees of freedom. The prospect of using high-power infrared lasers to do this created great excitement in the mid-1970s³. Initial reports seemed very promising. A CO₂ laser, with a power of ~20 MW cm⁻², was able to break a S–F bond in SF₆ (a 30- to 40-photon process). But in fact this was a thermal reaction: coherent pumping of energy could not compete³ with vibrational energy flow. The net effect of infrared laser pumping was incoherent heating, and the laser turned out to be an expensive Bunsen burner.

How could this redistribution be avoided? One group proposed decoupling a large molecule into segments using a heavy atom, in order to limit vibrational energy redistribution⁴. In recent years the focus of laser chemistry has shifted to the use of carefully crafted laser pulses with precise envelopes, timing and even phases to achieve selectivity and control^{2,5,6}. But so far, selective, non-thermal multiphoton laser chemistry has only worked in very small molecules with well-separated energy levels.

Another way to increase reaction rates would be if vibrational energy could somehow be channelled into a specific reactive site or bond. Ordinary random fluctuations may take a long time to get the energy to the right place. Some bias, such as an external field or energy gradient, could direct the energy flow^{7–9}. Alternatively, coherent migration of vibrational energy among the chromophores of a branched molecule could concentrate enough energy on a shorter timescale.

The new experiment¹ is based on this principle, but it turns out that the type of molecules used not only funnel the energy usefully, they stop it leaking out. Dendrimers are an interesting class of tree-like macromolecules with branched architecture^{10,11}.

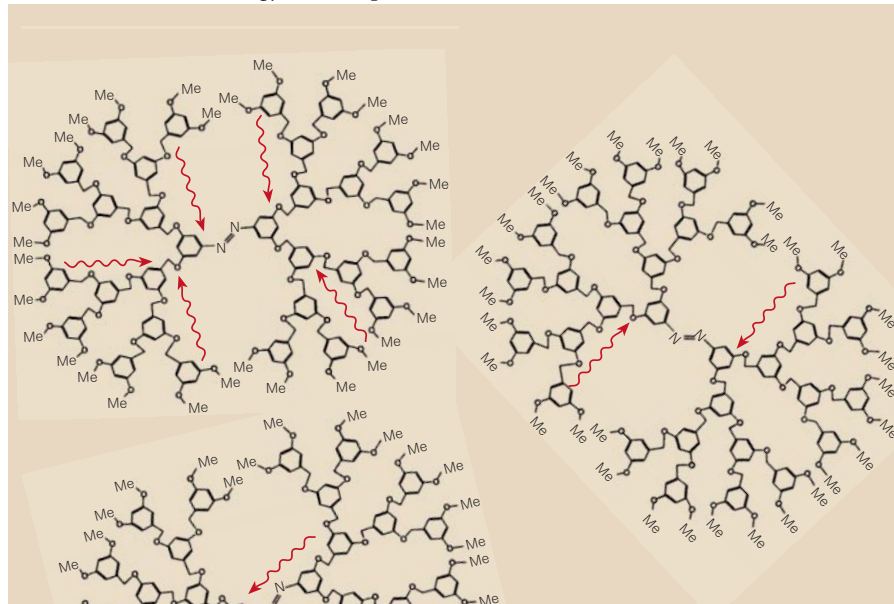


Figure 1 Three dendrimers in solution. Vibrational excitations are induced by an infrared laser, and funnelled towards the centre of each molecule. One contains five excitations — just enough energy in total to isomerize this molecule. Because they are prevented from leaking into the solvent by the close-packed, rigid outer shell, these excitations with eventually combine to cause the isomerization.

The dendrimeric backbone contains a large number of aromatic benzene rings, each of which can be excited into a stretching vibration by an infrared photon. With low light intensity, no one ring gets excited by many photons, but the absorption of a single photon by several rings puts enough energy in the dendrimer to make an isomerization possible. Intramolecular funnelling of the energy to the reaction site appears to be the rate-limiting factor. This is reminiscent of thermally activated unimolecular reactions at low pressures, where intermolecular activation determines the rate¹².

No attempt is made to coherently pump the energy, and the absorption of light is strictly linear. The role of the dendrimeric structure is twofold. First, it serves as a light-harvesting device that collects many excitations close to the reactive site. Second, it maintains these excitations for a long time by slowing the escape of vibrational energy to the surrounding medium. In a dendrimeric geometry the density of atoms increases rapidly with size (or branching generation) and the atoms at the boundary become closely packed, hindering the motions that lead to the loss of energy. The trapped excitations then have plenty of time to explore phase space (that is, to be randomly redistributed between bonds) and eventually combine in the desired region of the molecule where the reaction takes place.

This picture of an unconstrained interior protected by a rigid boundary is supported by nuclear magnetic resonance measurements. (The degree of rigidity may depend on the type of dendrimer and the solvent's quality, as shown by molecular dynamics simulations¹³.) Further evidence for this process is provided by the absence of reactivity when the reactive site is on the outside rather than the inside of the dendrimer. Moreover, the increase in reactivity is observed only in high-generation dendrimers, where the exterior is closely packed and therefore stiff.

The reported rate scales with the fifth power of light intensity. This can be explained as follows: if each dendrimer has N benzene rings and if a fraction s of the rings has been excited (on average, among many molecules), then the probability that a particular dendrimer will carry n vibrational quanta is the poissonian $P(n) = e^{(-Ns)} (Ns)^n / n!$. Because $s \ll 1$, only a small fraction of the molecules will acquire the necessary energy; but all these molecules will eventually react, as the energy does not dissipate. The fraction s is proportional to the light intensity and the s^n factor therefore implies n th power dependence on the light intensity. The azobenzene system needs five photons to induce the reaction, which explains the $n = 5$ observed dependence on incident power. (In the multiphoton laser experiments, the non-linear dependence on incident power reflects the nonlinear absorption of light; here the

absorption is strictly linear, and the non-linearity comes from the need to create many excitations in the same molecule.)

Similar funnelling of energy occurs naturally in photosynthesis, where assemblies of chlorophyll antennae absorb energy and direct it to a reaction centre, where a primary electron transfer takes place. Dendrimeric geometry is very attractive as a possible artificial antenna¹⁴ — the absorbing units can be closely packed without much support structure, and a simple chemical change of the central group may induce a large conformational rearrangement of the dendrimer¹⁵. So this new-found double role of dendrimers, as light-harvesting antennas as well as prisons for vibrational energy, may allow new forms of supramolecular photochemistry. □

Shaul Mukamel is in the Chemistry Department, University of Rochester, PO Box RC 270216, Rochester, New York 14627-0216, USA.

1. Jiang, D.-L. & Aida, T. *Nature* **388**, 454–456 (1997).
2. Rice, S. A. in *Proc. 20th Solvay Conf. Chemistry* (ed. Gaspard, P.) *Adv. Chem. Phys.* Vol. 101 (Wiley, New York, in the press).
3. Black, J. G., Yablonovitz, E., Bloembergen, N. & Mukamel, S. *Phys. Rev. Lett.* **38**, 1131–1134 (1977).
4. Lopez, V., Fairen, V., Lederman, S. M. & Marcus, R. A. *J. Chem. Phys.* **84**, 5494–5503 (1986).
5. Warren, W. S., Rabitz, H. & Dahleh, M. *Science* **259**, 1581–1589 (1993).
6. Mukamel, S. *Principles of Nonlinear Optical Spectroscopy* (Oxford Univ. Press, New York, 1995).
7. Kopelman, R. *et al. Phys. Rev. Lett.* **78**, 1239–1242 (1997).
8. Bar-Haim, A., Klafter, J. & Kopelman, R. *J. Am. Chem. Soc.* **119**, 6197–6198 (1997).
9. Devadoss, C., Bharathi, P. & Moore, J. S. *J. Am. Chem. Soc.* **118**, 9635–9644 (1996).
10. Newkome, G. R., Moorefield, C. N. & Vogtle, F. *Dendritic Molecules* (VCH, Weinheim, 1996).
11. Tomalia, D. A., Naylor, A. M. & Goddard, W. A. *Angew. Chem. Int. Edn Engl.* **29**, 138–175 (1990).
12. Robinson, P. J. & Holbrook, K. A. *Unimolecular Reactions* (Wiley, New York, 1972).
13. Murat, M. & Grest, G. S. *Macromolecules* **29**, 1278–1285 (1996).
14. Gust, D. *Nature* **386**, 21–22 (1997).
15. Junge, D. M. & McGrath, D. V. *Chem. Commun.* 857 (1997).

Neurobiology

Long-distance long-term depression

Roger A. Nicoll and Robert C. Malenka

Those who study learning and memory generally agree that many forms of these processes are due to long-lasting, use-dependent changes in the strength of synapses in the brain. Depending on the pattern of synaptic activation, synapses can show either a long-term potentiation (LTP) or a long-term depression (LTD). Both forms of plasticity are spatially restricted — that is, they seem to be limited only to those synapses in which the plasticity has been induced. This synapse-specificity has been used in neural-network models that store information, and it suggests that the signals responsible for the changes in synaptic strength are highly localized.

Now, an article by Fitzsimonds *et al.*¹ (page 439 of this issue) indicates that a form of LTD may spread over hitherto unimagined distances. Using simultaneous recordings from three interconnected neurons in culture, the authors show that LTD is not limited to the synapses at which it is induced, but that it spreads to other synapses in the microcircuit. They propose that this spread involves signals travelling retrogradely across the synapse and, in addition, signals that travel intracellularly for considerable distances. Moreover, although the changes are widespread, specific rules determine which synapses are affected (Fig. 1, overleaf).

The idea that LTD^{2–6} and LTP^{7–9} may not be strictly synapse-specific is not entirely new. So why is the paper by Fitzsimonds *et al.*¹ so provocative? First, it challenges the classic view — as originally put forward by Cajal — that the neuron is 'dynamically polarized'; that is, the information received by the dendrites is passed forward through

the axon to the synapses on the next cell. The newly discovered backward-propagation of LTD indicates that there is also an intracellular, retrograde, axon-terminal-to-dendritic flow of information (on a much slower time-scale). In this way, the input of the neuron is 'told' the state of the output of the neuron. So dendrites not only integrate synaptic potentials, but they also integrate retrograde signals from various axonal outputs of the cell. This has important implications for models of learning that are based on back-propagation algorithms¹⁰.

Second, the cell-biological mechanisms that must be proposed to account for many of the observations go far beyond those previously considered. Earlier work^{2–5,7–9} indicates that synaptic plasticity may spread laterally via a signal in the extracellular space. But Fitzsimonds *et al.* suggest that a retrograde signal (at synapse 2 in Fig. 1) generates a secondary intracellular signal in the presynaptic cell. This signal must travel all the way back into the dendrites (to depress synapse 1) and also throughout the presynaptic axonal arbour (to depress synapse 5). So this signal can depress synapses, whether it reaches the synapse from the presynaptic or postsynaptic side. The same, or a different, signal produced by excitatory or inhibitory synapses spreads LTD through the dendrites of the postsynaptic cell to other inputs that converge onto the cell (to depress synapse 4), yet it is incapable of forward propagation down the axon (to synapse 3). All in all, this is a level of complexity that few would have ever considered.

What remains to be done? As is the case with most provocative studies, far more questions are raised than answered. First, how far

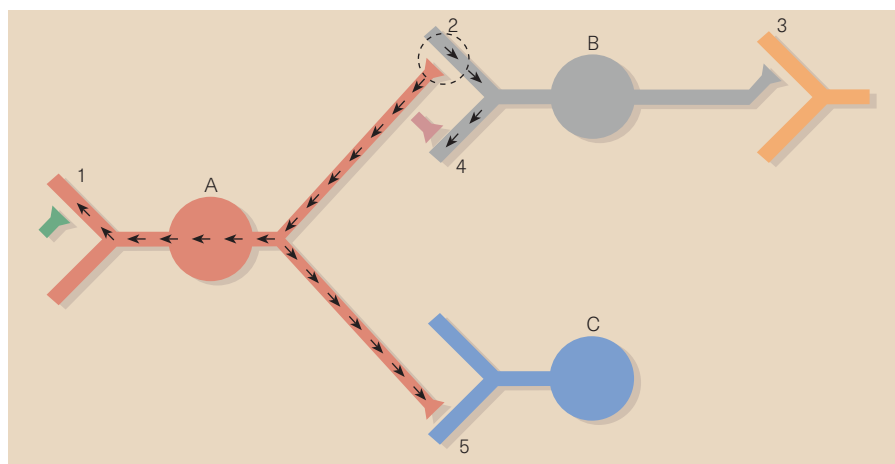


Figure 1 The spread of LTD (long-term depression) in a neural network in culture, as proposed by Fitzsimonds *et al.*¹. If LTD is induced at excitatory synapse 2 (circled synapse on cell B), LTD travels back through cell A to depress transmission at synapse 1. It also travels laterally within both the presynaptic cell (to synapse 5 on cell C) and the postsynaptic cell (to synapse 4 on cell B). If synapse 2 is inhibitory, LTD travels only laterally within the same postsynaptic cell (to synapse 4). Importantly, LTD induced at synapse 2 — whether it is excitatory or inhibitory — does not travel forward to affect synapse 3. Furthermore, the synapses that are affected by the propagated LTD can be either excitatory or inhibitory.

does the LTD spread? The authors provide a hint that synapses close to that in which the LTD is induced are subjected to more LTD than those further away. But whether there is a limit to the spread of LTD remains one of the key issues. Second, might LTD of a single synapse be enough to generate the propagating signal(s), or are a number of synapses needed to generate sufficient signal? The inputs in the study by Fitzsimonds *et al.* were large (hundreds of pA), indicating that each input consisted of a large number of synapses. This is in striking contrast to the intact nervous system, where single inputs generally make less than a dozen synapses. Third, is the propagated LTD graded or all-or-none? Fourth, because the propagated LTD requires the spread of a signal, a delay might be expected in its onset compared with the site of LTD induction — an issue that is not considered by Fitzsimonds *et al.* Fifth, what is the mechanism for the induction of LTD at inhibitory synapses? The LTD is induced while the membrane potential is voltage-clamped, so there is no obvious route for calcium entry, which is a normal requirement for plasticity. However, the LTD is clearly induced postsynaptically, because it is blocked if the membrane potential is held at hyperpolarized levels.

Finally, do the new findings in culture reflect what is going on in the intact nervous system? Many of the predictions have not been tested in the slice preparation because of the difficulty in identifying individual neurons in a circuit. One aspect that has been addressed by other investigators is the postsynaptic spread of LTD. When similar induction protocols are used, LTD seems to be limited to the activated synapses. In addition, many excitatory neurons in the central nervous system are projection neurons which send their axons great distances. This

anatomy is asking a lot of a back-propagation signal, yet if this mechanism is to be used by the brain it must find its way back to the dendrites. Lastly, there is little evidence that monosynaptic inhibitory inputs in the hippocampus show LTD or that they

transmit LTD to excitatory synapses.

The new work of Fitzsimonds *et al.* shouts out two very loud marching orders to scientists working in the synaptic-plasticity field. First, they need to find the underlying mechanisms and signals that are involved in this propagated synaptic plasticity in cell culture. Second will come the search for similar phenomena in the intact nervous system. Given the provocative nature of the present findings, we will not have to wait long for answers. □

Roger A. Nicoll is in the Departments of Cellular and Molecular Pharmacology, and Physiology, University of California, San Francisco, California 94143, USA. Robert C. Malenka is in the Departments of Psychiatry and Physiology, University of California, San Francisco, California 94143, USA.

1. Fitzsimonds, R. M., Song, H.-j. & Poo, M.-m. *Nature* **388**, 439–448 (1997).
2. Lynch, G. S., Dunwiddie, T. & Gribkoff, V. *Nature* **266**, 737–719 (1977).
3. Christie, B. R. & Abraham, W. C. *Synapse* **10**, 1–6 (1992).
4. Scanziani, M., Malenka, R. C. & Nicoll, R. A. *Nature* **380**, 446–450 (1996).
5. McMahon, L. L. & Kauer, J. A. *Neuron* **18**, 295–305 (1997).
6. Cash, S., Zucker, R. S. & Poo, M.-m. *Science* **272**, 998–1001 (1996).
7. Bonhoeffer, T., Staiger, V. & Aertsen, A. *Proc. Natl Acad. Sci. USA* **86**, 8113–8117 (1989).
8. Schuman, E. M. & Madison, D. V. *Science* **263**, 532–536 (1994).
9. Engert, F. & Bonhoeffer, T. *Nature* **388**, 279–284 (1997).
10. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Nature* **323**, 533–536 (1986).

V(D)J recombination

From RAGs to stitches

David T. Weaver and Fredrick W. Alt

DNA double-strand breaks (DSBs) can be formed by ionizing radiation and, if unrepaired, they are lethal to the cell. DSBs are also made as an initial step of the programmed V(D)J recombination pathway. This is used by lymphocytes to assemble the variable (V), diversity (D) and joining (J) gene segments which encode the variable regions of antigen receptors.

How do cells join double-strand breaks in their chromosomal DNA? Genetic evidence has shown that to rejoin the site-specific DSBs generated during V(D)J recombination, some of the proteins that are used in all cells to rejoin other DSBs are required. However, the mechanisms by which DSB-repair proteins effect the repair reaction is unknown. Now, papers by Ramsden *et al.*¹, Grawunder *et al.*² and Wilson *et al.*³ on pages 488, 492 and 495 of this issue, respectively, along with studies elsewhere^{4–6}, provide new insights into the joining phase of V(D)J recombination, and identify a related end-joining reaction in the yeast *Saccharomyces cerevisiae*.

The lymphoid-specific recombination-activating gene proteins 1 and 2 (RAG1 and RAG2) initiate V(D)J recombination by generating DSBs at the conserved recombination signal sequences (RSSs) that flank all antigen-

receptor variable-region gene segments. This initiation reaction has been carried out *in vitro*, and it has been shown to generate hairpin coding ends and blunt RSS ends⁷ (Fig. 1a). To complete the reaction the RSS ends are precisely joined, whereas the coding ends undergo further processing (for example, opening of the hairpin or the loss and gain of nucleotides) before they are joined (Fig. 1b). In a significant advance, Ramsden *et al.*¹ and others^{4,5} have now coupled the RAG1/RAG2-dependent initiation step with one or the other joining reactions in a cell-free system.

The most striking new finding is that if RAG1/RAG2 are retained after the initiation step, the formation of coding-junction products is stimulated. Coupling initiation to joining would allow a RAG1/RAG2 complex to direct the joining of appropriate ends. It is not known whether RAG1 and RAG2 physically interact with coding ends to promote joining. However, RAG1 and RAG2 are associated with RSS ends after cleavage⁸, and this may contribute to the difficulty in observing RSS joining *in vitro*¹⁵. Because the efficiency of V(D)J joining reactions seems to be quite low¹, other factors or altered conditions may be required. Further studies are needed to show how accurately the *in vitro* reactions mimic true

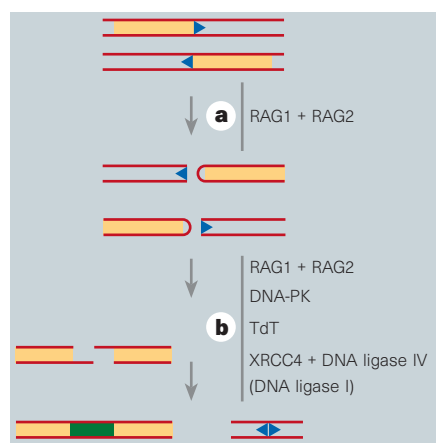


Figure 1 Ramsden *et al.*¹, Grawunder *et al.*² and Wilson *et al.*³ have studied the proteins involved in *V(D)J* recombination — the process by which the gene segments that encode the variable regions of antigen receptors are assembled. Double-stranded DNA breaks formed at sites of *V(D)J* recombination occur at the borders of coding DNA sequences (yellow) and recombination signal sequences (RSS; blue triangles). a, Cleavages mediated by the recombination-activating gene proteins RAG1 and RAG2 generate hairpin coding DNA ends and blunt RSS DNA ends. b, The joining phase of the reaction requires RAG1, RAG2, the components of the DNA-dependent protein kinase, DNA-PK (Ku70, Ku80 and DNA-PK_{cs}), XRCC4 and a DNA ligase. Terminal deoxynucleotidyl transferase (TdT), if present, can add untemplated nucleotides to joins.

V(D)J recombination — for example, by assaying for coupled coding and RSS joins (using inversional recombination substrates).

The RAG1/RAG2-mediated joining steps must act in coordination with the proteins that have been genetically implicated in *V(D)J* recombination — the Ku70, Ku80 and catalytic subunits of DNA-dependent protein kinase (DNA-PK), XRCC4 and terminal deoxynucleotidyl transferase (TdT). *In vitro* joining of the RSS⁴ and coding⁵ ends is inhibited by antibodies to Ku and DNA-PK_{cs}, respectively, indicating the direct involvement of Ku and DNA-PK in these reactions. But how Ku/DNA-PK_{cs} interact with RAG1/RAG2 synaptic complexes is not yet known.

The final step of joining requires the covalent ligation of DNA ends. Mammalian cells have three DNA ligases — DNA ligases I, II/III and IV. In contrast, *S. cerevisiae* seems to have two DNA ligases: CDC9, which is most closely related to mammalian DNA ligase I; and DNL4, discovered by Wilson *et al.*³, which is closest to DNA ligase IV. Unlike CDC9, DNL4 is not essential in yeast, but Wilson *et al.* find that *dnl4* mutants are deficient in non-homologous end joining — a repair mechanism in yeast that is similar to *V(D)J* joining in mammalian cells.

Ku has also been found to be involved in the non-homologous double-strand-end-joining pathway in yeast^{9,10}. By extension, this

suggests that DNA ligase IV might be used in *V(D)J* recombination. In support of this theory, DNA ligase IV associates with XRCC4 in mammalian cells^{2,6}, and Grawunder *et al.*² have found that XRCC4 stimulates the activity of ligase IV *in vitro*. XRCC4 binds to a non-catalytic, carboxy-terminal domain of DNA ligase IV, which contains the two BRCT motifs that are found in the breast cancer susceptibility gene 1 (*BRCA1*) and other DNA-repair-associated factors⁶. The *XRCC4* gene complements both coding and RSS *V(D)J* recombination defects, as well as the DSB repair defect of the XR-1 cell line¹¹. So ligase IV would be predicted to be used in all of these processes.

Ramsden *et al.*¹ found that when they added back DNA ligase I (but not ligases III and IV or bacteriophage T4 DNA ligase) to the *in vitro* reaction, *V(D)J* joining was stimulated, and resulted in a more random distribution of joins. This led them to suggest that ligase I plays a major role in *V(D)J* joining, in apparent conflict with the proposal by Grawunder *et al.*² that ligase IV is the main ligase for this process. But the observed lack of ligase IV activity might be explained if pre-formed XRCC4/DNA ligase IV complexes are needed to stimulate *in vitro* joining. The involvement of both ligases might reflect variations in DNA-end structures or accessibility. Another possibility is that the *in vitro* activity of ligase I may not be representative of its role *in vivo*. Further research is needed to determine whether DNA ligases I, IV or both, are used in *V(D)J* joining *in vivo*.

Mutations in XRCC4 and the components of DNA-PK have been shown to confer sensitivity to ionizing radiation — so, there is a striking parallel between *V(D)J* joining and chromosome-damage repair. However, it is not yet clear how the DSBs that initiate these processes are linked to the common repair steps that follow. Considering the requirement for RAG1/RAG2 in late stages of *V(D)J* joining, a key to end joining may be dictated by protein-protein contacts in these different contexts. The further development of *in vitro* *V(D)J* recombination and end-joining reactions should allow us to learn more about these processes and the proteins that are involved. □

David T. Weaver is at the Center for Blood Research, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. Frederick W. Alt is at the Howard Hughes Medical Institute, The Children's Hospital and Center for Blood Research, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

1. Ramsden, D. A., Paull, T. T. & Gellert, M. *Nature* **388**, 488–491 (1997).
2. Grawunder, U. *et al.* *Nature* **388**, 492–495 (1997).
3. Wilson, T. E., Grawunder, U. & Lieber, M. R. *Nature* **388**, 495–498 (1997).
4. Cortes, P. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 14008–14013 (1996).
5. Weis-Garcia, F. *et al.* *Mol. Cell. Biol.* (in the press).
6. Critchlow, S. E., Bowater, R. P. & Jackson, S. P. *Curr. Biol.* **7**, 588–598 (1997).
7. McBlane, J. F. *et al.* *Cell* **83**, 887–395 (1995).
8. Agrawal, A. & Shatz, D. *Cell* **89**, 43–53 (1997).
9. Milne, T. *et al.* *Mol. Cell. Biol.* **16**, 4197–4198 (1996).
10. Bolton, S. J. & Jackson, S. P. *EMBO J.* **15**, 5093–5103 (1996).
11. Li, Z. *et al.* *Cell* **83**, 1079–1089 (1995).

Daedalus

Clearing the air

Atmospheric fog is a serious traffic hazard. The resulting accidents tend to be blamed on moral lapses ('motorway madness') but are probably caused by misinterpretation of the greatly shortened lines of sight. Fog droplets absorb very little light, but scatter it confusingly. No lamp can help; its light is merely scattered into a bright blur.

Infrared imaging has often been advocated as a way out, with little success. During World War II, naive proposals of this sort were so common that one British defence researcher made a rubber stamp bearing the words "Infrared does not work!" Daedalus now hopes to make it work.

The effective scattering area of a suspended particle can be quite different from its physical cross-section. This is because its permittivity and refractive index can vary wildly in the region of an absorption or a resonance. In a suitable gap between absorptions the effective cross-section of a droplet can be under 10% of its physical one. Now, the refractive index of water falls pretty close to unity around 11 μm wavelength, a consequence of strong absorptions on either side. If it could be brought down exactly to unity, IR would go straight through each drop without reflection or deviation, though still with a little absorption. At this exact wavelength, fog would be transparent.

So DREADCO physicists are blundering around in artificial fog with narrow-band IR cameras, scanning their seeing wavelength in search of the 11- μm window. If it is clear enough, a car-mounted fog-penetrating IR camera should be feasible. If it is murky, DREADCO chemists will devise a 'fog additive' to dissolve in the droplets. Its absorption bands will bring the refractive index of water to exactly unity at one IR wavelength. Released along a motorway, possibly from vehicle exhausts themselves, it will create a transparent IR tunnel along which cars will bowl in perfect safety.

Meanwhile, DREADCO's astronomers are applying the converse reasoning to the elusive 'dark matter' in many galaxies. If it is interstellar dust, why doesn't it reveal itself by scattering starlight? By sheer bad luck, says Daedalus, it has a cross-section minimum in the optical region. But such a minimum must be compensated by a nearby maximum, such as the Fröhlich frequency, at which a particle's optical cross-section may be twenty times its physical one. Find the Fröhlich frequency of the dark matter, and it will blot out the heavens.

David Jones

Obituary

Martin Schwarzschild (1912–97)

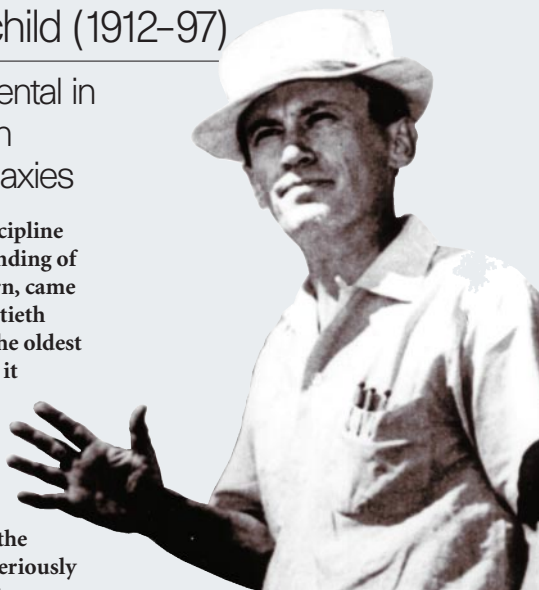
Astrophysicist instrumental in developing the modern theory of stars and galaxies

Astrophysics, as a theoretical discipline designed to provide an understanding of the observable Universe, was born, came of age and flourished in the twentieth century. Although it is perhaps the oldest quantitative science, before 1900 it was almost exclusively concerned with the positions and motions of planets and stars. For dynamical purposes, planets and stars could be treated as point objects. Only in the twentieth century did we begin seriously to inquire into the interior structure, origin and evolution of these ‘points’.

Martin Schwarzschild, who died on 30 April, was at the centre of this scientific transformation. He was born and educated in pre-war Germany, arriving in the United States (via Norway) in 1937. After a decade of military service and junior positions at Harvard and Columbia, he moved to Princeton as a professor. There he joined Lyman Spitzer Jr, the two of them establishing Princeton University Observatory as a centre of excellence in theoretical astrophysics, often working together, until they both died within weeks of each other almost exactly half a century from their initial appointments in the spring of 1947.

Nature and nurture conspired to provide an ideal environment for Schwarzschild. His uncle, Robert Emden (a professor of physics and meteorology), published the famous book *Gaskugeln* (*Gas Spheres*) in 1907, establishing the treatment of stars as gaseous bodies within which the forces of Newton’s gravity and gas pressure are in equilibrium. His father, Karl Schwarzschild, was the author of the famous Schwarzschild (spherical) solution to Einstein’s equations. Martin became probably the leading developer of the theory of stellar evolution.

Largely because of his contributions to this field, we can now follow the progress of a typical star after it contracts out of the interstellar medium and begins to burn its nuclear fuel through a sequence of phases: main sequence (like the Sun) to red giant (like Antares) to degenerate dwarf (like Sirius B). By now there are innumerable quantitative tests of the theory of stellar evolution. And it works! One must consider it the greatest triumph of astrophysical theory in this century that



despite a few remaining puzzles (the nature of type-II supernovae, and the low level of neutrino emission from the Sun, for example), there is an extraordinarily good match between theoretical models of stars — singly, in pairs and in clusters — and the bulk of the astronomical evidence.

Schwarzschild’s style as a scientist combined traits rarely found together in a single person: a broad and imaginative creativity based on a thorough understanding of physical principles; a passion for detailed and precise numerical work to determine the consequences of physical models; and an enormous respect for observations — the facts of nature as the ultimate arbiters of truth.

A few examples make these points vividly. Most early work on stellar structure had treated stars as being of uniform composition. Schwarzschild realized that, through the expected processes of nuclear burning, stars would develop discontinuities in chemical composition, smoothed out only by relatively inefficient diffusion. In a series of papers with many associates (including Alan Sandage and Fred Hoyle) that began in the late 1940s and continued through the early 1970s, he showed that these discontinuities have powerful effects on stellar structure, producing surprising internal effects such as burning in unstable shells, or externally observed effects such as extended red-giant envelopes. Approximate analytic theory, or perturbation theory applied to existing exact analytic solutions, could not begin to address this problem even qualitatively, because the equations of stellar structure are highly nonlinear and have two boundary conditions, at the centre and the surface. So it was necessary to develop new

numerical algorithms and even new methods of computation. As a result, Schwarzschild worked with John von Neumann in the 1940s at Princeton to develop the first digital electronic computers used for scientific research.

A second example is taken from the dynamics of galaxies, to which field Schwarzschild returned after his formal retirement in 1979. He had realized that the usual assumption concerning the symmetry of galaxies — that they are essentially symmetrical around an axis of rotation — was just that: an untested assumption. He then began a series of studies to examine the issue observationally and to see whether self-consistent models could be constructed that would deviate from this assumption, yet agree with observations. The work again involved massive numerical computations, and it revolutionized our understanding of ‘elliptical’ galaxies. They are now known to be triaxial systems in general, for which rotation is not the principal dynamical force.

Schwarzschild’s capacity to strip all problems to their essential elements, combined with his vivacious personality, produced one of the great teachers and expositors of his discipline. He was equally famous for large public lectures and for his intimate ties to and guidance of his students. A person of enormous modesty and charm, he was as well loved by peers and colleagues as by students and others in all occupations whom he met in daily life.

The recipient of essentially every major honour available to a theoretical astrophysicist, he was awarded the US National Medal of Science, unfortunately posthumously on 30 April 1997, with the citation: “For his seminal contributions to the theory of the evolution of stars, and his creative insights into galactic dynamics which form the basis of much of the contemporary astrophysics; and his lifetime of dedication to students. His influence on US astronomy in the second half of this century is unsurpassed.” A thoroughly cosmopolitan scientist, he was as honoured in the United Kingdom as in the United States. He won the gold medal of the Royal Astronomical Society in 1969 and was elected a foreign member of the Royal Society in 1996.

It seems unlikely that history will provide another such moment, when one individual can have such a large effect on our view of the essential elements of the Universe.

Jeremiah P. Ostriker

Jeremiah P. Ostriker is at the Office of the Provost, 3 Nassau Hall, Princeton University, Princeton, New Jersey 08544, USA.

Light-induced amphiphilic surfaces

The ability to control the surface wettability of solid substrates is important in many situations. Here we report the photogeneration of a highly amphiphilic (both hydrophilic and oleophilic) titanium dioxide surface. The unique character of this surface is ascribed to the microstructured composition of hydrophilic and oleophilic phases, produced by ultraviolet irradiation. The result is a TiO_2 -coated glass which is antifogging and self-cleaning.

We prepared a thin TiO_2 polycrystalline film from anatase sol on a glass substrate, which was annealed at 773 K. The film showed a water-contact angle of $72^\circ \pm 1^\circ$ before ultraviolet irradiation (Fig. 1a). After irradiation, water droplets spread out on the film, resulting in a contact angle of $0^\circ \pm 1^\circ$ (Fig. 1b). This change in wettability was clearer when a TiO_2 -coated glass was exposed to water vapour. Without ultraviolet irradiation the glass fogged (Fig. 1c), but on irradiation the glass became transparent (Fig. 1d), a remarkable antifogging effect.

We also measured the contact angle using oily liquids (such as glycerol trioleate and hexadecane). We found distinct contact angles for the TiO_2 film under normal conditions, but all the liquids spread across the surface on ultraviolet irradiation, with contact angles of $0^\circ \pm 1^\circ$. Irradiation created a surface that was highly hydrophilic and highly oleophilic. We observed the wettability change on both anatase and rutile TiO_2 surfaces of polycrystals or single crystals, independent of their photocatalytic activities. Even after storage in the dark for a few days, the high amphiphilicity of the TiO_2 surface was maintained. A longer storage period induced a gradual increase in the water-contact angle, revealing a surface wettability trend towards hydrophobicity. However, high amphiphilicity was repeatedly regenerated by ultraviolet irradiation.

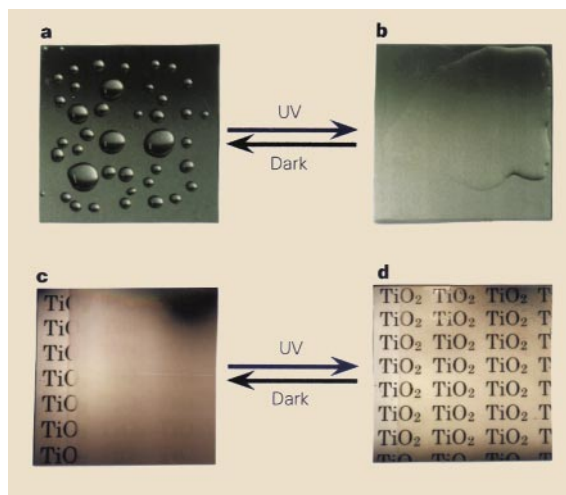
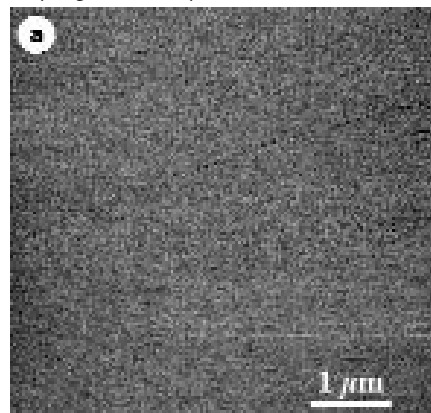


Figure 1 **a**, A hydrophobic surface before ultraviolet irradiation. **b**, A highly hydrophilic surface on ultraviolet irradiation. **c**, Exposure of a hydrophobic TiO_2 -coated glass to water vapour. The formation of fog (small water droplets) hindered the view of the text on paper placed behind the glass. **d**, Creation by ultraviolet irradiation of an antifogging surface. The high hydrophilicity prevents the formation of water droplets, making the text clearly visible.

The contact angle gives information about the macroscopic surface wettability. To gain information about surface wettability at a microscopic level we used friction force microscopy (FFM)¹. A rutile TiO_2 (110) single crystal was used to make these measurements because of its flat surface, a requirement for FFM. Before ultraviolet irradiation (Fig. 2a) we observed no difference in contrast, indicating microscopically homogeneous wettability. After irradiation, (Fig. 2b) hydrophilic (bright) and oleophilic (dark) areas of size 30–80 nm were clearly seen. Higher resolution images (Fig. 2b inset) show hydrophilic domains with regular rectangular shapes, aligned along the [001] direction of the (110) crystal surface, irrespective of the scanning direction. A gradual reversion to a lack of contrast was observed during the storage of the film in the dark. We conclude that the nanoscale separation between the hydrophilic and the oleophilic phases accounts for the highly amphiphilic character of the TiO_2 surface.

In combination with Fourier transform

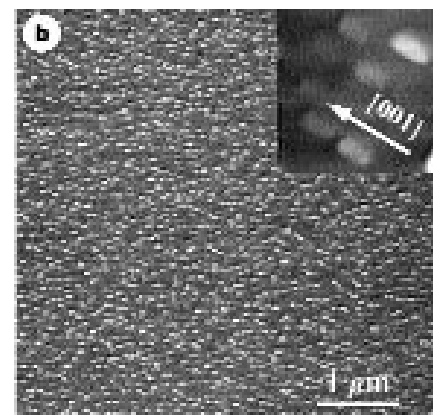


Figure 2 FFM images of a rutile TiO_2 (110) single crystal surface. **a**, A $5 \times 5 \mu\text{m}^2$ image before ultraviolet irradiation. **b**, The same surface after irradiation. Inset, topographic image ($245 \times 245 \text{ nm}$) acquired by rotating the sample stage through 45° to the large-scale image. The tip of the Si_3N_4 cantilever is hydrophilic, so hydrophilic areas are bright and hydrophobic areas are dark.

infrared spectroscopy and X-ray photoelectron spectroscopy studies, we propose the following model. Ultraviolet irradiation may create surface oxygen vacancies at bridging sites, resulting in the conversion of relevant Ti^{4+} sites to Ti^{3+} sites² which are favourable for dissociative water adsorption^{3,4}. These defects presumably influence the affinity to chemisorbed water of their surrounding sites, forming hydrophilic domains, whereas the rest of the surface remains oleophilic, as seen by FFM (Fig. 2b). The rectangular hydrophilic domains are areas where dissociative water is adsorbed, associated with oxygen vacancies that are preferentially photogenerated along the [001] direction of the (110) plane; the same direction in which oxygen bridging sites align⁵. Because a liquid droplet is substantially larger than the hydrophilic (or oleophilic) domain, it instantaneously spreads on such a surface, thereby resembling a two-dimensional capillary phenomenon. Microscopically, the hydrophilic and oleophilic areas are distinguishable, but macroscopically, the TiO_2 surface exhibits high amphiphilicity. On long-term storage in the dark, the chemisorbed hydroxyl groups are replaced with oxygen from the air³.

Such highly amphiphilic surfaces have many practical applications. Besides the antifogging effect (Fig. 1d), they are self-cleaning. Various substrates coated with TiO_2 films, regardless of their photocatalytic activities, showed obviously cleaner surfaces than those without TiO_2 coating after being hung outdoors for six months. Ultraviolet irradiation from sunlight is sufficient to maintain the amphiphilic surface, so that hydrophilic or oleophilic contaminants on the surfaces are easily removed by rain. This unique amphiphilic surface should be superior to a monohydrophilic or monoleophilic surface. The present system is independent of conventional photocatalytic reactions⁶, but the combination of these

two surface characters give TiO₂ materials a promising future.

Rong Wang, Kazuhito Hashimoto

Akira Fujishima

Department of Applied Chemistry,
Faculty of Engineering, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
e-mail: akira-fu@chem.t.u-tokyo.ac.jp

Makoto Chikuni, Eiichi Kojima

Atsushi Kitamura, Mitsuhide Shimohigoshi
Toshiya Watanabe

Research and Development Centre,
TOTO Ltd, 2-8-1, Honson Chigasaki-shi,
Kanagawa 253, Japan

1. Wilbur, J. L., Biebuyck, H. A., MacDonald, J. C. & Whitesides, G. M. *Langmuir* **11**, 825–831 (1995).
2. Shultz, A. N. *et al. Surface Sci.* **339**, 114–124 (1995).
3. Hugenschmidt, M. B., Gamble, L. & Campbell, C. T. *Surface Sci.* **302**, 329–340 (1994).
4. Henderson, M. A. *Surface Sci.* **355**, 151–166 (1996).
5. Murray, P. W., Condon, N. G. & Thornton, G. *Phys. Rev. B* **51**, 10989–10997 (1995).
6. Fujishima, A. & Honda, K. *Nature* **238**, 37–38 (1972).

S-nitrosylation regulates apoptosis

Nitric oxide (NO) modulates the biological activity of proteins by direct interactions with their iron centres. It can also S-nitrosylate cysteines to form S-nitrosothiols. Such reactions affect the activity of membrane-bound, cytosolic and nuclear proteins including the NMDA receptor¹, haemoglobin² and transcription factors such as NF- κ B³ and OxyR. NO is potentially toxic, inducing both apoptosis and necrosis. Here we show that NO-mediated S-nitrosylation of the cysteine-containing enzymes that mediate apoptosis (caspases and tissue-transglutaminase, tTG) regulates the balance between apoptosis and necrosis.

Apoptosis is triggered by caspases⁴ cleaving substrates such as poly(ADP ribose)-polymerase (PARP)⁵, and by tTG crosslinking of proteins⁶. We studied the effects of NO donors on cell death, PARP proteolysis and tTG activity. At 1 mM concentration, S-nitroso-N-acetylpenicillamine (SNAP) releases 60 nM free NO (ref. 3), therefore millimolar concentrations of pharmacological NO donors are needed to produce low micromolar concentrations of free NO in cells.

SNAP blocks apoptotic death in Jurkat T cells exposed to anti-CD95 (APO-1, Fas) antibody in a dose-dependent manner, as shown by FACS analysis (Fig. 1a), Hoechst 33342 staining and DNA laddering (not shown). The reduction of apoptosis was detectable after 3 h and significant after 6 h ($P < 0.001$).

Conversely, increasing concentrations of SNAP progressively increased necrotic cell death, as shown by trypan blue uptake (Fig.

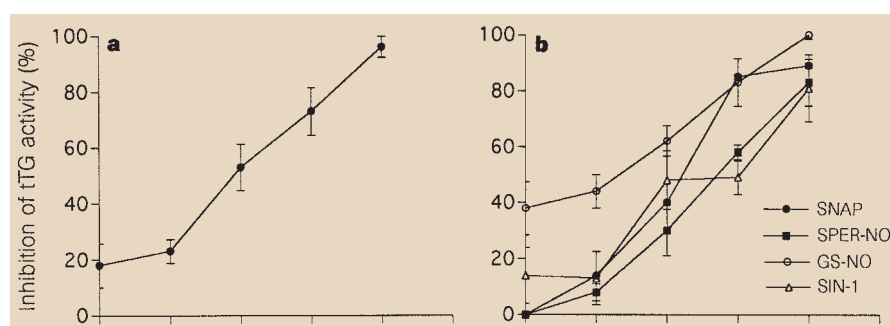
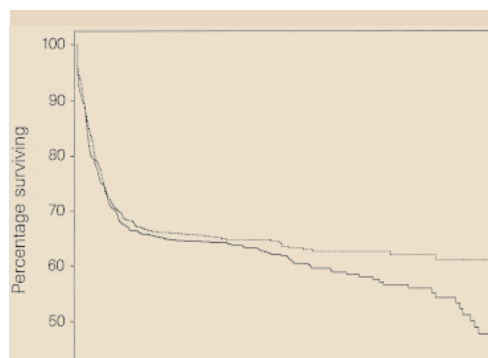


Figure 1 Mean dose- and time-dependent effect of SNAP on cell death assessed by: **a**, flow cytometry using propidium iodide⁶; and **b**, trypan blue uptake. Cells were exposed to SNAP (0.1–10 mM) alone or with 1 μ g ml⁻¹ anti-CD95 antibody and evaluated after 6 h. **c**, Inhibition of anti-CD95-induced proteolytic cleavage of PARP by SNAP. Cells were untreated (lane 1), incubated with SNAP (2–5) or 1 μ g ml⁻¹ anti-CD95 antibody (6–9) and with the NO donor (7–9) for 90 min. Extracts were analysed by western blotting using the monoclonal anti-PARP antibody C-2-10.



1b) and electron microscopy (not shown). Six hours after CD95 ligation, lactate dehydrogenase release increased by 10, 37, 45 and 180 IU l⁻¹ using 0, 0.1, 1.0 and 10 mM SNAP, respectively (baseline activity was 79 IU l⁻¹). Similar apoptosis/necrosis regulation was obtained by application of SNAP to Jurkat cells exposed to 30 μ M C6-ceramide (not shown).

SNAP (Fig. 1c) or GS-NO (S-nitroso-glutathione; not shown) inhibited CD95-triggered digestion of PARP by caspase-3 in a dose-dependent manner. This precedes the shift from apoptosis to necrosis triggered by CD95 ligation (90 min as opposed to 6 h). SNAP also inhibited the activity of purified tTG (Fig. 2a). This time-independent inhibition was prevented by NO scavengers. In the absence of D/L-dithiothreitol (DTT) the baseline activity of tTG was reduced from 1,200 $\pm$ 160 to 41 $\pm$ 6 nmol [³H]putrescine

per h per mg of protein, and significantly more NO was required to produce the same level of inhibition. DTT reduces the cysteine in the active site of tTG, activating the enzyme. The strong shift of the dose-response curve in the absence of DTT indicates that NO might react with this same cysteine.

Molecular modelling shows that nine of the 20 cysteines in tTG are close to the surface. NO reacts with one of the three titratable -SH groups: in fact, only two residues were detectable after exposure to 1 mM SNAP or 0.1 mM GS-NO. Similar results have been found for caspase-3 (ref. 7). Moreover, 1 mM spermine-NO eliminated two of the three titratable -SH groups, suggesting that free spermine, a natural tTG-substrate, binds to the substrate site and that this differs from the NO-bound active site.

Jurkat cells express little tTG, which is not required in CD95-induced apoptosis,

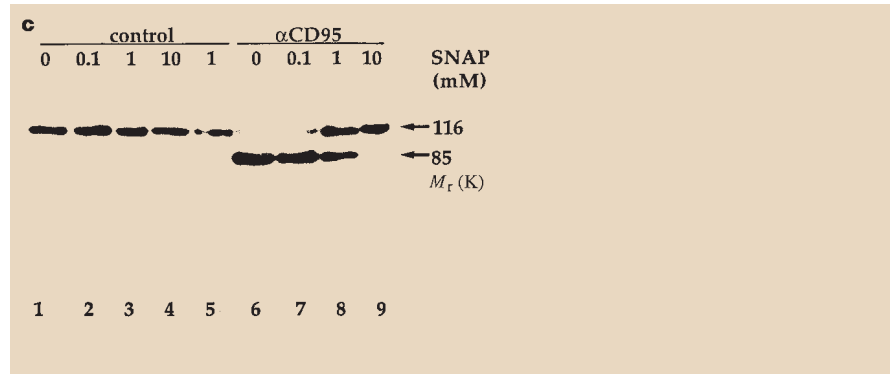


Figure 2 a, Effect of SNAP on purified tTG activity. SNAP was added to the enzyme preparation just before initiating the reaction with [³H]putrescine and *N,N*-dimethylcasein. Purified tTG (86 nM) was incubated with 30 mM DTT for 20 min (ref. 6). Means of three experiments. **b**, Inhibition of tTG activity in SK-N-BE(2) cell extracts by increasing concentrations of the NO donors SNAP, spermine-NO, GS-NO or 3-morpholinosydnonimine (SIN-1). Extracts were exposed to donors immediately before the assay and incubated with 30 mM DTT for 20 min. Baseline activity was 215 $\pm$ 24 and 161 $\pm$ 18 pmol [³H]putrescine incorporated into *N,N*-dimethylcasein per h per mg protein in the presence and absence of DTT, respectively. Means of three experiments.

but SK-N-BE(2) human neuroblastoma cells require tTG for apoptosis⁶ and die by necrosis on exposure to NO. In these cells NO donors inactivated intracellular tTG in a dose-dependent manner (Fig. 2b). Whereas NO donors, such as SNAP, release limited free NO (3–4 log units lower than donor concentration)³, GS–NO is an NO-carrier. Indeed, the GS–NO inhibition curve is shifted by 3 log units.

NO can elicit either apoptosis or necrosis in neuronal cells, although there is still some confusion as to the definition of these types of cell death. Whether a cell will undergo apoptosis or necrosis is regulated by the intensity of the stimulus⁸. Necrosis predominates in response to acute exposure to relatively high NO concentrations when NO can react with O_2^- to generate high levels of $ONOO^-$ causing neuronal damage¹. In contrast, prolonged exposure to low doses of NO induces apoptosis. NO and $ONOO^-$ inhibit respiration at the level of cytochrome *c* oxidase and complexes I–III, respectively⁹. Thus, mitochondrial function has been suggested to be a factor in determining the balance between apoptosis and necrosis by NO⁸. Recently, mitochondria have been suggested to be important in apoptosis¹⁰. Cytochrome *c* is released from mitochondria and activates caspases¹¹, although a possible interaction of NO with its haem remains to be demonstrated.

We propose an additional, or alternative, mechanism to explain the cellular decision to enter the necrotic or apoptotic pathway. NO can S-nitrosylate the active-site cysteine residues in at least two important effector elements of apoptosis. We have shown that NO converts apoptosis into necrosis in CD95-ligated Jurkat cells, inactivates tTG by S-nitrosylation and inactivates caspase-mediated PARP degradation. This inhibition acts both *in vitro* and intracellularly. In our model, relatively high concentrations of NO are needed to cause significant inactivation of tTG and PARP cleavage, suggesting that a relatively intense NO stimulus is necessary to block apoptosis and allow free radical-mediated necrotic damage to take place.

It is not clear whether the reversion of apoptosis into necrosis by NO can be generalized. It occurs to different degrees in Jurkat cells ligated by CD95 or treated with C6-ceramide and in neuroblastoma cells with retinoids. Caspases 1 and 3 have recently been shown to be inhibited by NO, resulting in the inhibition of TNF α -induced apoptosis⁷. NO also inhibits the DNA-binding activity of both NF- κ B³ and AP-1¹², two cysteine-transcription factors implicated in apoptosis^{13,14}.

S-nitrosylation may simultaneously inactivate several parts of the apoptotic machinery and may balance apoptosis and necrosis. The distinct role of mitochondria (cytochromes) and S-nitrosylation in determin-

ing which NO-induced pathway of cell death is followed must be evaluated in each individual cell-death model to clarify mechanisms and modulators such as glutathione.

Gerry Melino, Francesca Bernassola

Biochemistry Laboratory,
Istituto Dermopatico Immacolata,
University of Rome Tor Vergata, Rome 00133, Italy
and Department of Biology, Coppito,
University of L'Aquila, 67100 L'Aquila, Italy
e-mail: melino@utovrm.it

Richard A. Knight

Cystic Fibrosis Department,
National Heart and Lung Institute,
Imperial College, London SW3 6LR, UK

**Maria Tiziana Corasaniti, Giuseppe Nisticò
Alessandro Finazzi-Agrò**

Department of Experimental Medicine,
University of Rome Tor Vergata, 00133 Rome, Italy

1. Lipton, S. A. *et al.* *Nature* **364**, 626–632 (1993).
2. Jia, L., Bonaventura, C., Bonaventura, J. & Stamler, J. S. *Nature* **380**, 221–226 (1996).
3. Matthews, J. R. *et al.* *Nucleic Acids Res.* **24**, 2236–2242 (1996).
4. Kumar, S. & Lavin, M. F. *Cell Death Differ.* **3**, 255–267 (1996).
5. Nicholson, D. W. *et al.* *Nature* **376**, 37–43 (1995).
6. Melino, G. *et al.* *Mol. Cell. Biol.* **14**, 6584–6596 (1994).
7. Dimmler, S., Haendeler, J., Nehls, M. & Zeiher, A. M. *J. Exp. Med.* **185**, 601–607 (1997).
8. Ankarcona, M. *et al.* *Neuron* **15**, 961–973 (1995).
9. Lizasoain, I., Moro, M. A., Knowles, R. G., Darley-Usmar, V. & Moncada, S. *Biochem. J.* **314**, 877–880 (1996).
10. Kroemer, G., Petit, P., Zamzami, N., Vayssiere, J.-L. & Mignotte, B. *FASEB J.* **9**, 1277–1287 (1995).
11. Kluck, R. M., Bossy-Wetzel, E., Green, D. R. & Newmeyer, D. D. *Science* **275**, 1132–1136 (1997).
12. Tabuchi, A., Sano, K., Tsuchiya, T. & Tsuda, M. *FEBS Lett.* **351**, 123–127 (1994).
13. Lipton, S. A. *Nature Med.* **3**, 20–22 (1997).
14. Smeyne, R. J. *et al.* *Nature* **363**, 166–169 (1993).

Thermal-expansion materials not so new

Robert Cahn in News and Views¹ discusses work by Sleight's group^{2,3} demonstrating that the phase ZrW_2O_8 has a negative thermal expansion. But the conclusion is not novel: this phase is one of many known for decades to exhibit the phenomenon. (Three relevant reviews^{4–6} were not cited in the News and Views article.) Further, Sleight's group and Cahn both fail to note the discovery of the very significant, incredibly large NZP–CTP ($NaZr_2P_3O_{12}$ or $Ca_{0.5}Ti_2P_3O_{12}$) structural family first found by us⁷, subsequently studied by many groups and now commercialized. (Individual references to other relevant publications can be found in refs 4–6. A bibliography containing 100 articles on the NZP family, including references to the examples mentioned below, is available on request from the authors.)

The NZP–CTP family is essentially infinite (including solid solutions), with members going from very negative coefficient of thermal expansion ($NaTi_2P_3O_{12}$ is one of the most negative with α of -5.5×10^{-6} per °C) to fairly positive. This material is

obviously of great interest because the coefficient of thermal expansion (CTE) can be tailored so easily. Moreover, some of these phases melt only above 1,700 °C (ZrW_2O_8 is known to 'melt' or dissociate at ~800 °C and has received so little attention partly for that reason).

We also wish to comment on two other points raised^{2,3}. Our first point is the claim that the crystallographic anisotropy of CTE in the eucryptite, spodumene and cordierite family "limits the value" of such materials. But the opposite is the case. F. A. Hummel discovered and patented the negative CTE in eucryptite in 1948. One of us (R. R.) worked out the detailed parent phase diagram for $Li_2O-Al_2O_3-SiO_2$ in the same year, and our group has been using it widely since. Stookey used this material in the Corning 'glass-ceramic' process, one of the most significant breakthroughs in materials processing in 50 years. Millions of casserole dishes, saucepans and entire stove-tops consisting of these phases are now made each year, and Schott is using this family to produce possibly the largest materials monoliths ever made in the form of 10 m $\times$ 1-m-thick transparent telescope blanks.

Our second point concerns the crystallographic explanation for the phenomenon, another long-studied area. In the NZP family, the exactly analogous high-temperature X-ray analysis has been reported (up to 1,000 °C), and the thermal expansion anisotropy and negative expansion in the NZP structure shown to result from the coupled rotational motion of ZrO_6 octahedra and PO_4 tetrahedra. A similar explanation was given by Sleight *et al.*^{2,3} for the negative expansion of the ZrW_2O_8 phase, although in more detail and at lower temperatures. But, more significantly, we were able to reduce not only α to zero, but the very anisotropy of α essentially to 'zero' in the system $(Ca,Sr)Zr_4P_{16}O_{24}$. The "possible" composites suggested by Sleight's group have already been made: real composites of thermodynamically co-stable negative- α NZP structure phases and positive- α YIG ($Y_3Fe_2Fe_3O_{12}$, yttrium iron garnet) were painstakingly developed to make ferrimagnetic zero-expansion ceramics for radar-invisible space-mirrors that would not distort with variation of exposure to the Sun.

We believe that the points we raise here illustrate why a knowledge of the older literature is valuable. Not only can time and money be saved by removing the need to repeat what has already been done, but older papers can contain gems of ideas which, in the light of new knowledge and equipment, can sparkle anew and indicate fresh directions.

Rustum Roy, Dinesh Agrawal

Materials Research Laboratory,
Pennsylvania State University,
University Park, Pennsylvania 16802, USA
e-mail: web@alpha.mrl.psu.edu

1. Cahn, R. W. *Nature* **386**, 22–23 (1997).
2. Evans, J. S., Mary, T. A., Vogt, T., Subramanian, M. A. & Sleight, A. W. *Chem. Mater.* **8**, 2809–2823 (1996).
3. Mary, T. A., Evans, J. S. O., Vogt, T. & Sleight, A. W. *Science* **272**, 90–92 (1996).
4. Chu, C. N., Saka, N. & Suh, N. P. *Mater. Sci. Eng.* **95**, 303 (1987).
5. Roy, R., Agrawal, D. K. & McKinstry, H. A. *Annu. Rev. Mater. Sci.* **19**, 59–81 (1989).
6. Agrawal, D. K. *J. Mater. Educ.* **16**, 139 (1994).
7. Alamo, J. & Roy, R. *J. Am. Ceram. Soc.* **67**, C78 (1984).

Season of birth predicts mortality in rural Gambia

We present evidence that events in early life strongly influence the adult survival prospects of rural Africans. Our analysis of births and deaths in three Gambian villages dating back to 1949 shows that people born during the annual 'hungry season' are up to 10 times more likely to die prematurely in young adulthood. A permanent effect of malnutrition on the development of the immune system during fetal growth seems a likely explanation.

Patterns of nutrition and disease in rural communities from sub-Saharan Africa are heavily influenced by a clear divide between wet and dry seasons. In the Gambia, the wet season (July–October) coincides with an annual hungry period when staple foods from the previous harvest are seriously depleted¹. This is compounded in adults by an intensive agricultural workload, and in young children by diarrhoeal disease and malaria which peaks during the rains².

These stresses combine to cause intra-uterine growth retardation³, severe growth faltering in infancy and early childhood², and weight loss in adults¹ during this season. Impaired fetal growth reduces birth-weight by roughly 200–300 g and doubles the incidence of low-birthweight babies (< 2.5 kg), but can be reversed by supplementing maternal diet³. This seasonality has been documented for over 40 years⁴, the intensity varying according to rainfall, disease patterns and crop yields.

Since 1949 detailed records of births and deaths have been collected in three rural

subsistence-farming villages (Keneba, Kantong Kunda and Manduar)⁴. These records provide data on month of birth for 3,102 individuals, born between 1949 and 1994, for whom current fate (number of deaths, 1,077, and date of death) is known with certainty up to the end of 1994.

Analysis of variance in this data set revealed highly significant month-of-birth effects with highest mortality in births from July to December, two months longer than the usual hungry season divide. At an early age, deaths were similar in groups born between January and June and between July and December, but from the age of 15 those born in the hungry season had a greater mortality (Fig. 1).

When examined for deaths at all ages the season-of-birth effect was not significant (hungry, 581/1,627; harvest, 496/1,475; log-rank $\chi^2 = 2.3$, $P = 0.13$), but for those surviving to the age of 15 the odds ratio for premature death was 3.65 (38/460 against 10/415; log-rank $\chi^2 = 15.3$, $P < 0.0001$). This increased consistently to reach 10.4 for deaths past the age of 35 (9/68 against 1/69; logrank $\chi^2 = 6.2$, $P = 0.013$). Classification of the 48 deaths occurring after the age of 15 (based on clinic records or verbal autopsies) revealed: 17 infections, eight maternal deaths (at least two infection-related), three cancers, four renal failures, three epilepsy-related, five accidents and three other causes, with five unknown.

These results indicate that prenatal or early postnatal events affect the future health of rural Gambians in a manner first manifested around puberty and amplified with increasing age. The linkage between early life events and later viability has clear parallels with the 'infant and fetal origins of disease' hypothesis proposed by Barker for chronic diseases of affluence⁵. However, in the present findings, mortality was dominated by infections and pregnancy-related deaths with none being related to chronic degenerative disease.

Nutritional programming during fetal life seems the most likely causal factor (although early infections can have permanent effects on immune function⁶). Obser-

vations on the acute effects of protein-energy malnutrition suggest a likely involvement of cell-mediated immunity⁷. As early as 1810 Menkell linked malnutrition with lymphoid-tissue atrophy (especially of the thymus), and 'nutritional thymectomy' became a common medical term⁸. Several components of the human immune system mature early in fetal life⁹; deficits in organ growth and development occurring *in utero* are more serious and long-lasting than those caused by later malnutrition¹⁰. Low-birthweight babies may have sustained impairment of immune competence as infants and children⁹. In animals the effects of fetal undernutrition on immunological function have even been recorded in F₂ and F₃ offspring¹¹.

The early, *in utero*, sensitivity of the immune system to malnutrition would fit with our current observation that babies born up to two to three months after the peak of the hungry season remain vulnerable suggesting that fetal growth impairment in mid-pregnancy may have a carry-over effect even when maternal nutrition improves in late pregnancy. Likewise there is an intriguing similarity between the deviation of the Kaplan–Meier survival plots, the natural onset of thymic involution at puberty, and the well-documented post-pubertal decline in immune function.

Our current finding has emerged from a unique setting combining a seasonal experiment of nature with detailed longitudinal demographic records, but the general principle of fetal programming of immune function is likely to be of broad significance in other populations, especially those with marginal nutrition and high levels of exposure to disease.

Sophie E. Moore

Timothy J. Cole

Elizabeth M. E. Poskitt

Bakary J. Sonko

Roger G. Whitehead

Ian A. McGregor

Andrew M. Prentice

MRC Dunn Nutrition Unit, Hills Road,

Cambridge CB2 2DH, UK

and MRC Dunn Nutrition Unit, PO Box 273,

Fajara, Banjul, The Gambia, West Africa

e-mail: andrew.prentice@mrc-dunn.cam.ac.uk

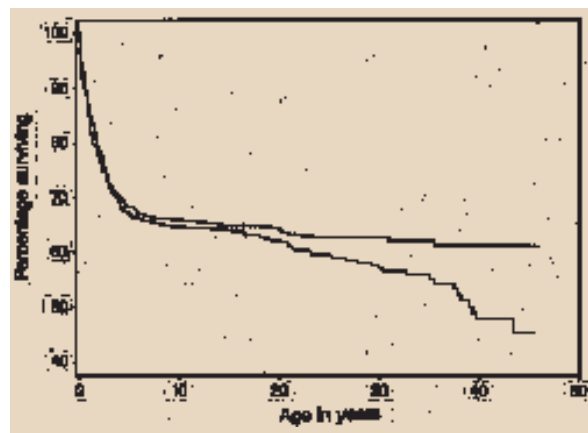


Figure 1 Kaplan–Meier survival plots for rural Gambians divided by season of birth. Harvest season, dotted line; hungry season, solid line.

1. Prentice, A. M., Whitehead, R. G., Roberts, S. B. & Paul, A. A. *Am. J. Clin. Nutr.* **34**, 2790–2799 (1981).
2. Cole, T. J. in *Seasonality and Human Ecology* (eds Ulijaszek, S. J. & Strickland, S. S.) 89–106 (Cambridge Univ. Press, 1993).
3. Ceesay, S. M. *et al. Brit. Med. J.* (in the press).
4. Billewicz, W. Z. & McGregor, I. A. *J. Biosoc. Sci.* **13**, 219–240 (1981).
5. Barker, D. J. P. *Mothers, Babies and Diseases in Later Life* (BMJ Publishing Group, London, 1994).
6. Aaby, P., Andersen, M. & Knudsen, K. *Int. J. Epidemiol.* **22**, 156–162 (1993).
7. Chandra, R. K. *Proc. Nutr. Soc.* **52**, 77–84 (1993).
8. Beisel, W. R. *J. Nutr.* **122**, 591–596 (1992).
9. Chandra, R. K. *Lancet* **ii**, 1393–1394 (1974).
10. Winnick, M. *Pediatrics* **47**, 969–978 (1971).
11. Beach, R. S., Gershwin, M. E. & Hurley, L. S. *Science* **218**, 469–471 (1982).

Rise and eclipse of an elusive planet

In Search of Planet Vulcan: The Ghost in Newton's Clockwork Universe

by Richard Baum and William Sheehan
Plenum: 1997. Pp. 310. \$28.95, £17.55

Joseph N. Tatarewicz

Most people know the planet Vulcan as the home of *Star Trek's* Mr Spock and his kin rather than as the intra-Mercurial planet that came — and went — not quite a century ago. Thanks to Richard Baum and William Sheehan, accomplished researchers and writers who have separately brought the history of astronomy to a wide audience in previous works, interested readers can now find a similar interplay of logic and passion in the story of the historical Vulcan.

Although the details of the search for this planet, purported to lie within the orbit of Mercury, have been expounded in more esoteric works, never have they been brought together in such a comprehensive, elegant and accessible way. This book is an excellent account of the episode, its historical context and its colourful personalities, as well as a fine introduction and guide to the detailed historical scholarship on which it is based.

Isaac Newton tamed the motions of the planets with his gravitational theory, but he left many residual problems. Two generations of observational and mathematical astronomers then extended and refined Newton's mathematical techniques to resolve these problems and handle newly discovered objects and the peculiarities of their motions.

During the eighteenth and nineteenth centuries, theorists faced with ill-behaved objects often despairingly toyed with the idea of modifying Newton's essential theory. Each time, they discovered a new technique or observation that turned anomaly into confirmation of Newtonian theory.

In the most famous of these episodes, the planet Uranus (discovered by William Herschel after Newton) stubbornly refused to follow the most sophisticated implementations of Newtonian theory.

On the basis of these deviations, John Couch Adams in England and Urbain Le Verrier in France independently calculated the likely orbit of a hypothetical trans-Uranian planet capable of perturbing Uranus to the degree required. The discovery of Neptune in 1846 from these predictions rescued Newtonian gravitational theory, and supplied a model that would nurture searches for other planets, including those that led to the discovery of Pluto in 1930.

The model infused the young Le Verrier

with confidence that such an approach could rescue Newtonian theory from what was now its most serious problem — Mercury's orbit. Whereas a fictional ideal planet would orbit its sun in a precise ellipse, the mutual interactions of many real bodies produce changes in the various orbital elements.

Mercury's perihelion, it was well known, was advancing too fast. The extra 43 seconds of arc per century was the most glaring problem of gravitational astronomy, and the one to which Le Verrier turned as his career took off.

To a person with a hammer, the saying goes, everything looks like a nail. To Le Verrier, the hammer was the method that had worked so well with Neptune, and he became convinced that a planet, Vulcan, was perturbing Mercury's motion.

Ultimately, the problem would be solved by changing tools, and adopting the then almost unthinkable option of modifying Newtonian theory. Einstein's general theory of relativity in 1915 neatly resolved Mercury's motion, but only at the expense of abandoning the Newtonian understanding of gravitation.

During the final half-century of purely Newtonian gravitational astronomy, Vulcan was widely believed to exist, relentlessly and painstakingly stalked, and even discovered more than once. It was, as Baum and Sheehan aptly call it, the ghost in Newton's clockwork universe.

The mathematics of celestial mechanics is laborious, complex and esoteric, even for straightforward situations. Calculating 'backwards', from deviations to the charac-



Le Verrier: enthusiastic Vulcanologist.

ter of a putative perturbing body, is even more difficult. Once one has a likely location to search for the planet, the observational confirmation is equally difficult. In the case of Neptune (and Pluto), observers had to search a vast expanse of celestial real estate, looking for a tiny, dim object amid a myriad of others in a sky only imperfectly mapped and catalogued.

The hunters of Vulcan had a more circumscribed field (just within Mercury's orbit), but had to contend with the Sun. After difficult twilight searching had proved fruitless, they had to resort to the few frantic and fleeting moments of solar eclipses when the sky near the Sun becomes dark enough to reveal fainter objects.

The authors draw on their own considerable practical knowledge of astronomy, as well as the participants' own published and unpublished accounts, to give the reader a feel for such difficult practical matters. They also ably convey the adventure of such eclipse expeditions, which took observers to remote and sometimes dangerous locations.

The authors tell a marvellously engaging story, bringing a host of colourful characters to life. They allude to the surrounding social and political context in a way that helps the reader to place events in a long story that unfolds over nearly three centuries. Their explanations of the complex mathematical theories and techniques are models of clarity and economy.

The book is studded with helpful tables and illustrations, including portraits, astronomical photographs and diagrams. The extensive notes, bibliography and chronology nicely complement the narrative.

The Vulcan episode itself was largely concentrated in the second half of the nineteenth century, but the authors provide context and relevance by extending their treatment back to Newton and forward to modern times.

The caricature of the Vulcan episode has traditionally been that scientists looked in vain for Vulcan and, having exhaustively searched, eventually declared it nonexistent.

Baum and Sheehan demonstrate that the story is much more subtle and nuanced than that. By recounting subsequent and even recent discoveries of dim, Sun-grazing comets, asteroids and other small bodies, they remind us that even our comfortable inner Solar System is far from thoroughly explored. □

Joseph N. Tatarewicz is in the Department of History, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250, USA.
e-mail: tatarewicz@umbc.edu

Home is the hunter

Fairweather Eden

by Michael Pitts and Mark Roberts
Century: 1997. Pp. 356. £17.99

Asako Saegusa

Boxgrove in West Sussex, England, has over the years turned up many flint tools and fossil animal bones, supplying intriguing evidence that our 'primitive' ancestors hunted, killed and butchered large animals rather than scavenging for food. Although these findings themselves were compelling enough to spark heated debate among prehistorians, it was not until the discovery three years ago of a hominid fossil shin-bone that this Palaeolithic

site gained worldwide recognition (see *Nature* 369, 311–313; 1994).

The tibia belonged to a hominid from the Middle Pleistocene who stood six feet tall and weighed 13 stone (182 pounds or 83 kg). He was later identified by the archaeologists as *Homo heidelbergensis*, an ancestor of modern Europeans. The fossil was originally estimated to be 200,000 years old, but biostratigraphic evidence from the excavation site dated it to 500,000 years ago — the oldest hominid bone then to have been found in Europe.

Boxgrove man's status as the oldest European has since been challenged by discoveries at Dmanisi, Georgia, and at Atapuerca in northern Spain. Finds at Boxgrove have nevertheless provided us with a better understanding of Archeulian industry in Britain and the

behaviour of our ancestors. *Fairweather Eden* is a colourful insider's account of the 10-year excavation at Boxgrove, which started as a small-scale project when Mark Roberts, the site director, was an undergraduate at the Institute of Archaeology in London.

The book describes the progress of projects at Boxgrove, provides an historical perspective of European archaeological excavation, advances theories of hominid tool use and cognitive development and presents a snapshot of life in Britain half a million years ago.

So what do we know about the flint tools? Archeulian hand axes were previously described by some archaeologists as a by-product of tool manufacture. But bones of large animals such as extinct rhinos found at Boxgrove reveal cut marks made by flint tools, and some show filleting marks matching those of modern butchery; in fact, when a local butcher was asked to dismember a roe with reconstructed flint tools, he found them surprisingly effective. Also, the edges of the hand axes show wear that is reckoned could have been caused only by butchering meat.

In addition, circular perforations in an excavated horse shoulder are believed to have been caused by blows from a spear. This evidence, together with the recent discovery of 400,000-year-old wooden throwing-spears in Schöningen, Germany (see *Nature* 385, 807–810; 1997), indicates that Boxgrove hominids were robust in build and were big-game hunters.

Fairweather Eden is a well-documented account of recent discoveries that pieces together compelling and intelligent ideas about the evolution of early humans. □

Asako Saegusa has been appointed Tokyo correspondent of *Nature*.



In-flight meals

A frog who would a-wooing go attracts unwelcome attention. The fringe-lipped bat *Trachops cirrhosus* (below) fastens on to the mating call of frogs to find its daily diet. The yellow epauletted bat *Sturnira lilium* (left) prefers fruit, while other bats eat insects, fish and even scorpions. But only three out of more than 900 species of bat are vampires, and their record of biting humans is stronger in fiction than in fact. In *Bats in Question* (Smithsonian Institution Press, \$49 (hbk), \$24.95 (pbk)), mammal specialist Don E. Wilson attempts to rehabilitate a maligned group of animals, aided by stunning photography by Merlin D. Tuttle, giving a comprehensive guide to the different species and a host of bat facts.



Messages in melody

Musical Languages

by Joseph P. Swain

Norton: 1997. Pp. 384. \$32.50, £25. To be published in the UK in November

Christopher Longuet-Higgins

"How is music like language and so what if it is?" is Joseph Swain's truculent summary of his book. He sees himself as taking up where Deryck Cooke and Leonard Bernstein left off in their famous attempts to invigorate the analogy between music and language — attempts that were fiercely scorned at a time when the concept of structure was sweeping aside the ancient problems of expression and communication.

The topics that Swain selects for his comparative study include syntax, semantics, phonology, language acquisition, language evolution, artificial languages and language-using communities. Their musical counter-

parts have all come under fire in times of revolutionary change such as the early decades of the twentieth century.

Syntax, semantics and phonology deal, respectively, with the forms of utterances, their meanings and their realization in speech or writing. In traditional music, note order is as crucial as word order is in English. When we listen to a tune, our expectations and their satisfaction are governed by syntactic considerations, although we may not know what they are.

The psychological reality of musical syntax is demonstrated by our ability to identify wrong notes in music we have never heard before. Its value to the listener lies in establishing the tensions and resolutions that could never arise if the unfolding of the music were totally unpredictable — or totally predictable.

The semantics of music is something of a disaster area (but so, for that matter, is natural language semantics). Although a pastoral theme may well suggest the countryside, it cannot possibly assert that you are about to tread in a cowpat. But a ray of hope illuminates the attempt to assign meanings to vocal compositions, where the words and the music may be expected to tell the same story.

Swain invokes this principle to explain the dramatic effect of the chorus "For unto us a Son is born" in Handel's *Messiah*; but he could have done without it in discussing the music of the preceding aria, "The people that walked in darkness", where a lonely melodic line wanders about unpredictably in a minor key.

The phonology of a musical language is properly concerned with the conceptual (phonemic) relations between the notes rather than their acoustic (phonetic) realization. The tonal relations of Western classical music, so economically represented by stave notation, remained basically unchanged during the seventeenth, eighteenth and nineteenth centuries.

The real revolution wrought by the serialists at the beginning of the twentieth century was not to declare all the notes of the keyboard equal, but to throw out the conceptual distinction between homophonic intervals such as the major third and the diminished fourth, along with the bath water of decadent romanticism. But not for long: experiments by Jamshed Barucha have shown that the tonal baby has a way of crawling back into the musical reflexes of even quite fluent modernistic composers.

To illustrate his ideas, Swain discusses in detail such landmark compositions as Palestrina's "Sicut Cervis" and Beethoven's "Violin Concerto". But of more consequence for general musicology are his observations on musical evolution, artificial languages and the supporting role of musical communities. If he is right, the artistic health of languages such as jazz and rock seems assured; the music of small minorities, such as soli-

tary composers of electronic music, faces bleaker prospects.

The least convincing chapter is that on musical metaphor. Perhaps the next edition will replace it by one on prosody, a common concern of poets and songwriters.

Musical Languages is an untidy, enthusiastic, perceptive book, to be heartily recommended to all thoughtful musicians. □

Christopher Longuet-Higgins is in the Department of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Born to science

Words, Thoughts and Theories

by Alison Gopnik and Andrew N. Meltzoff
MIT Press: 1997. Pp. 268. \$30, £25.50

Peter Bryant

Anyone who has had any prolonged contact with children knows that their intellectual effectiveness changes remarkably as they grow older. This was first systematically demonstrated by Binet's intelligence test which showed ordinary children fumbling hopelessly with elementary verbal or spatial problems one year and solving them unhesitatingly the next. Later, many developmental psychologists, most notably Piaget, charted the details, many still controversial, of the kinds of problems young children can and cannot solve and how children get better at finding solutions as they grow older.

Intelligence testers, Piaget and most other developmental psychologists demonstrated that these changes tend to tail off around late adolescence (although they agreed that there is room for slight improvement during early adulthood) and they assumed that this endpoint in intellectual development is an intellectual summit. Until a child gets there, he or she is intellectually incomplete: after reaching the sum-



Early learning: scientific reasoning starts young, say Gopnik and Meltzoff.

mit, there will be no further improvements in intellectual machinery.

Psychologists vary greatly in how they characterize this endpoint, but Piaget's view probably remains the most influential. He claimed that the climax of intellectual change was scientific thinking: when children can form hypotheses, and test them systematically by looking at the effects of one variable at a time while holding the rest constant, they have reached the endpoint.

In their stimulating and well-written book, Alison Gopnik and Andrew N. Meltzoff attempt to overturn these usually unquestioned assumptions about the developmental endpoint and the late arrival of scientific reasoning.

Their main idea is that scientific reasoning is the cause, not the product, of chil-

New in paperback

Emotional Intelligence: Why It Can Matter More Than IQ

by Daniel Goleman

Bantam, \$13.95

"Goleman has written an interesting if somewhat naive book. ... Some will find the journalistic tone repugnant, but anyone interested in emotion is likely to discover challenging new ideas in this book: it should be read by social workers and agony aunts everywhere." Stuart Sutherland, *Nature* 379, 34 (1996).

The Power to Harm: Mind, Medicine and Murder on Trial

by John Cornwell

Penguin, £7.99

In 1989, Joseph Wesbecker shot 20 of his former

colleagues and then himself. He had been on Prozac, and his victims and their families sued its makers, Eli Lilly, for compensation. Cornwell uses the trial to warn of the potential apocalypse created by "the growing crisis over reductionist solutions to individual suffering and social disorder". Reviewed by Keri K. Gould in *Nature* 383, 783 (1996).

Reality Rules: Picturing the World in Mathematics. Vol. 1 — The Fundamentals. Vol. 2 — The Frontier

by John L. Casti

Wiley, each \$32.50, £19.99

"Casti Tours offers the most spectacular vistas of modern applied mathematics." Karl Sigmund, *Nature* 362, 669 (1993).

dren's intellectual development. Children test hypotheses about the physical and social world around them from the time they are born, and form new hypotheses when the old ones prove unsatisfactory. They do so exactly like scientists, according to Gopnik and Meltzoff.

In particular, the authors claim that the hypotheses children entertain are always falsifiable and that children treat them as such. Like good scientists, children recognize when one of their ideas has not worked out and they look for a new and better one.

And the endpoint? Well, all that happens is that after a while people get tired of the whole process and are content to stick with the hypotheses they happen to have reached by then. So the authors contend that adulthood is the end, not the beginning, of scientific reasoning for most people.

This argument goes flat against the idea of development as a series of changes from an immature to a more-or-less complete system. On the contrary, the authors argue that the child's intellectual machinery undergoes no change at all. The process of testing and forming new hypotheses is the same in adolescence as in the first few days of life.

The child's hypotheses get closer to the truth as time goes by and the result is that the child's knowledge about the world keeps on growing. But the intellectual machinery that leads to this growth is always the same, until it burns out in early adulthood.

The book deals mainly with children up to the age of two years and is mostly concerned with their understanding of physical objects and social interactions. There are sections on children's success in solving simple means-ends problems (such as getting a necklace into a jar), on the way in which they categorize objects and on the first stages of language acquisition.

Much of the argument is based on a reanalysis of well-known phenomena. For example, the authors dwell at length on the familiar observation (originally Piaget's) of the 'A not B' error: babies who see an adult hiding an object in one place (A) and then later in another place (B) will go on looking for the object at A even when they have just closely watched the adult put it in B.

According to the authors, babies make this particular mistake because they have formed a hypothesis that "the object will be where it appeared before" and arrive at this hypothesis on the basis of experiences such as dropping a toy out of their cot and not seeing it again until it has been tidied up and put back where it was originally.

This particular argument is typical of the book. It is ingenious and amusing — and it could be right. But the authors offer no evidence for it. They do throw doubt (quite rightly, in my view) on one competing theory — that the error is due to the

limitations in the children's ability to remember the sequence of events. Yet they fail to test, or even consider how to test, their own hypothesis even though it is, it seems to me, eminently testable.

Given the authors' willingness to offer a new view of babies' intellectual achievements, their unwillingness to find a way of testing it is ironic — after all, the book's main purpose is to persuade us how well young children set about testing their own hypotheses. The nearest they come to offering a direct test is in their account at the end of the book of their own work on language acquisition and on children's success in solving means-ends problems, in categorization tasks where they have to work out what has happened to an object they can no longer see ("object permanence").

Their research reveals specificity in the connections between different aspects of language acquisition and children's performance in different cognitive tasks. For example, they have shown a strong relationship between children's use of words denoting success and failure and their ability

to solve means-ends problems and a similarly close connection between their learning of words to do with appearance and disappearance and their success in object-permanence problems. No such connection was found across these two categories, however.

This is good research with genuinely interesting results, but it does not get near to establishing that these changes happen because babies test and abandon unsuccessful hypotheses and then form new ones.

We are left with a lively and entertaining book that will make us think again about children's intellectual development but will in the end convince no one. The authors will have to find some direct way of investigating how children form and abandon hypotheses if they want their fellow developmental psychologists to give up on one view and to espouse a completely different one. □

Peter Bryant is in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.

At a glance

Excellent ★★★★★ Good ★★★★ Fair ★★★ Poor ★

Quantum Mechanics in Chemistry

by Jack Simons and Jeff Nichols
Oxford University Press: 1977. Pp. 612.
£48.50, \$70

Good books on quantum mechanics in chemistry are always welcome. This book contains the standard wavefunction and orbital material with many examples. Angular momentum theory, determinants, configuration interaction, molecular rotation and vibration, and time-dependent processes (including a good introduction to scattering) are all clearly introduced. There is a section on quantitative methods, introducing the Hartree-Fock method and basis sets. The authors' expertise in more advanced methodology such as MCSCF and response theories appears in the last chapter. There is also a 125-page appendix.

This well-written text provides a good basis to standard quantum chemistry. But I have an important criticism. In the past four years, density functional theory has 'taken over' computational chemistry. Here it is given four pages only, written from a historical, out-of-date perspective. The theory should be given prominence in modern texts on quantum chemistry.

Nicholas C. Handy Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.

Range	★★★★
Depth	★★★★
Accuracy	★★★★
Up-to-dateness	★★
Accessibility	★★★★
Style	★★★★

Vertebrate Palaeontology, second edition

by Michael J. Benton
Chapman and Hall: 1997. Pp. 452. £75, \$110 (hbk); £24.99, \$50 (pbk)

Aiming to outline more than 500 taxonomic groups of living and fossil vertebrates in fewer than 400 pages of well-illustrated text is ambitious. Michael J. Benton has successfully updated and expanded the first edition (1990) to make it a more thorough introduction to the topic. For example, the new developments in the understanding of tetrapod and avian evolution are here, along with new illustrations, references from 1996, detailed discussion 'boxes' on topics such as the "birds of Las Hoyas" and "Permian burrowing dicynodonts of the Karoo", and new cladograms.

Sections on preservation, ecology, biomechanics, plate tectonics and theories of extinction are all necessary to help explain and interpret the peculiarly biased nature of the fossil record, especially to biologists without a palaeontological background. But these days many, if not most, students of vertebrate fossils have little idea about basic vertebrate zoology. They will need to supplement this book with a standard text on vertebrate zoology.

Douglas Palmer 31 Mawson Road, Cambridge CB1 2DZ, UK.

Range	★★★★
Depth	★★★★
Accuracy	★★★★
Up-to-dateness	★★★★
Accessibility	★★★★
Style	★★★★

Propagation of activity-dependent synaptic depression in simple neural networks

Reiko Maki Fitzsimonds, Hong-jun Song & Mu-ming Poo

Department of Biology, University of California at San Diego, La Jolla, California 92093-0357, USA

Triple whole-cell recordings from simple networks of cultured hippocampal neurons show that induction of long-term depression at glutamatergic synapses is accompanied by a back propagation of depression to input synapses on the dendrite of the presynaptic neuron. The depression also propagates laterally to divergent outputs of the presynaptic neuron and to convergent inputs on the postsynaptic neuron. There is no forward propagation of depression to the output of the postsynaptic neuron and no presynaptic propagation accompanying long-term depression at GABAergic synapses. Activity-induced synaptic modification is therefore not restricted to the activated synapse, but selectively propagates throughout the neural network.

The development and plasticity of the nervous system involve activity-dependent modification of synaptic connections^{1–3}. Long-term potentiation (LTP) and long-term depression (LTD) of central synapses induced by repetitive activity in various regions of the brain have been implicated as the cellular mechanisms that underlie learning and memory^{4–7}. These modifications are known to be input-specific, so only activated pathways are affected. But within the vicinity of activated pathways, other non-activated synapses also become modified, leading to more distributed synaptic modifications within a neural network. For example, LTP generated at one synaptic input to a CA1 hippocampal neuron was found to spread to adjacent synapses on the same or on a different postsynaptic neuron^{8–11} and to result in LTD in more distant neurons^{12,13}. Such spread of synaptic modification appears to depend on membrane-permeable or secreted factors released by the activated synapse, and the extent of influence on other synapses depends on their physical proximity. On the other hand, the spread of synaptic modification can also be mediated by signalling within the neuron. On the postsynaptic side, LTP at one synaptic input can result in heterosynaptic modification of other convergent inputs^{14–19}. On the presynaptic side, changes may spread through the cytoplasm to other divergent outputs. In *Xenopus* nerve-muscle cultures, LTD induced at one neuromuscular synapse can propagate to other synapses made by the same neuron onto other myocytes, apparently by signalling within the neuronal cytoplasm²⁰. A similar presynaptic cytoplasmic mechanism may also account for the spread of short-term depression of inhibitory synapses observed in cerebellar slices²¹.

Here we find that induction of LTD at glutamatergic synapses within small networks of cultured hippocampal neurons is accompanied by a retrograde spread of depression to synapses on the dendrites of the presynaptic neurons (back propagation). The depression also spreads laterally to synapses made by divergent outputs of the presynaptic neuron (presynaptic lateral propagation) or to convergent inputs on the postsynaptic neurons (postsynaptic lateral propagation). In contrast, there was no forward propagation of depression to output synapses of the postsynaptic neuron. Moreover, we found no evidence for back propagation or presynaptic lateral propagation of depression resulting from LTD induced at GABAergic synapses. Taken together, these results indicate that there is an extensive yet selective distribution of activity-induced synaptic changes within the neural network. Furthermore, the discovery of back propagation of LTD indicates

that there is an axon-dendritic flow of information which conveys signals from the outputs of a neuron to its inputs. It gives added credence to the idea that the back-propagation algorithm, which was used effectively for supervised learning in artificial neural networks^{22,23}, may be implemented in simple biological neural networks.

Triplet networks of hippocampal neurons

Low-density cultures of hippocampal neurons were prepared from embryonic day (E) 18–20 rat embryos²⁴ and used for experiments 10–14 days after plating. For the present study we searched for isolated groups of three interconnected hippocampal neurons (triplets) located in the microscopic field of view (a diameter of ~950 μm) in the absence of other neurons (Fig. 1a). The connectivity in the triplet was examined by simultaneous whole-cell perforated patch recording from the soma of all three neurons. For each neuron, we recorded glutamatergic (excitatory) postsynaptic currents (EPSCs) or GABAergic (inhibitory) postsynaptic currents (IPSCs) in response to brief step-depolarizations initiated in each neuron under voltage-clamp conditions ($V_c = -80 \text{ mV}$). By examining the latency of onset, the time course and the reversal potential of synaptic currents, all monosynaptic connections among three neurons in the isolated triplet and autaptic connections made by each neuron onto itself were characterized. As shown in Fig. 1b, c, the latency of onset of monosynaptic EPSCs and IPSCs following a 1-ms presynaptic depolarization had a mean value in the range of 1.5 to 2.6 ms, whereas polysynaptic responses exhibited much longer latencies ($>5 \text{ ms}$). We have thus defined the monosynaptic connection by the presence of postsynaptic currents that exhibit a latency $\leq 2.6 \text{ ms}$ and a monotonic rising phase. The monosynaptic connections used here had a mean latency of 1.8 ± 0.4 and $1.5 \pm 0.4 \text{ ms}$ ($n = 80$ each) for EPSCs and IPSCs, respectively (see also Fig. 1d, e). For an unambiguous interpretation, experiments were done only on triplets with limited reciprocal synaptic connections. Many neurons in the triplets form autapses, which share physiological properties with synapses^{25–27}. For simplicity of description, we used the terms ‘synapse’ and ‘autapse’ to refer to the ensemble of all synaptic and autaptic contacts (or boutons) made by a neuron.

Induction of LTD

We first characterized the induction of LTD at both glutamatergic and GABAergic synapses in these hippocampal cultures. Pairs of

Table 1 Summary of propagation of synaptic depression in small neural networks

Network configuration	Site of LTD induction		Site of propagation		N†	Propagation
	EPSC/IPSC	Depression (%)*	EPSC/IPSC	Depression (%)*		
	(1) Back propagation					
	EPSC	43.5 ± 4.9‡	EPSC	27.6 ± 3.0‡	12	Yes
	EPSC	49.8 ± 7.0‡	IPSC	48.4 ± 7.5‡	5	Yes
	IPSC	48.3 ± 7.6‡	EPSC	4.7 ± 2.3	4	No
	IPSC	52.2 ± 11.5‡	IPSC	-6.0 ± 10.2	5	No
	Forward propagation					
	EPSC	40.6 ± 4.9‡	EPSC	4.5 ± 3.1	3	No
	EPSC	24.8 ± 3.5‡	IPSC	-8.3 ± 7.4	4	No
	(2) Presynaptic lateral propagation					
	EPSC	43.8 ± 6.3‡	EPSC	41.9 ± 7.0‡	10	Yes
	IPSC	51.2 ± 5.3‡	IPSC	3.7 ± 1.9	6	No
	(3) Postsynaptic lateral propagation					
	EPSC	46.1 ± 6.9‡	EPSC	29.5 ± 5.9‡	6	Yes
	EPSC	44.3 ± 15.0‡	IPSC	49.3 ± 17.2‡	3	Yes
	IPSC	32.7 ± 5.9‡	EPSC	20.7 ± 4.2‡	3	Yes
	IPSC	56.7 ± 12.5‡	IPSC	19.2 ± 9.4‡	3	Yes

* The value of depression represents the percentage reduction in the amplitude of EPSCs and IPSCs during the first 10–15 min following the induction of LTD.

†N, total number of triplets examined for each condition.

‡Significantly different from zero ($P < 0.01$, Mann-Whitney test).

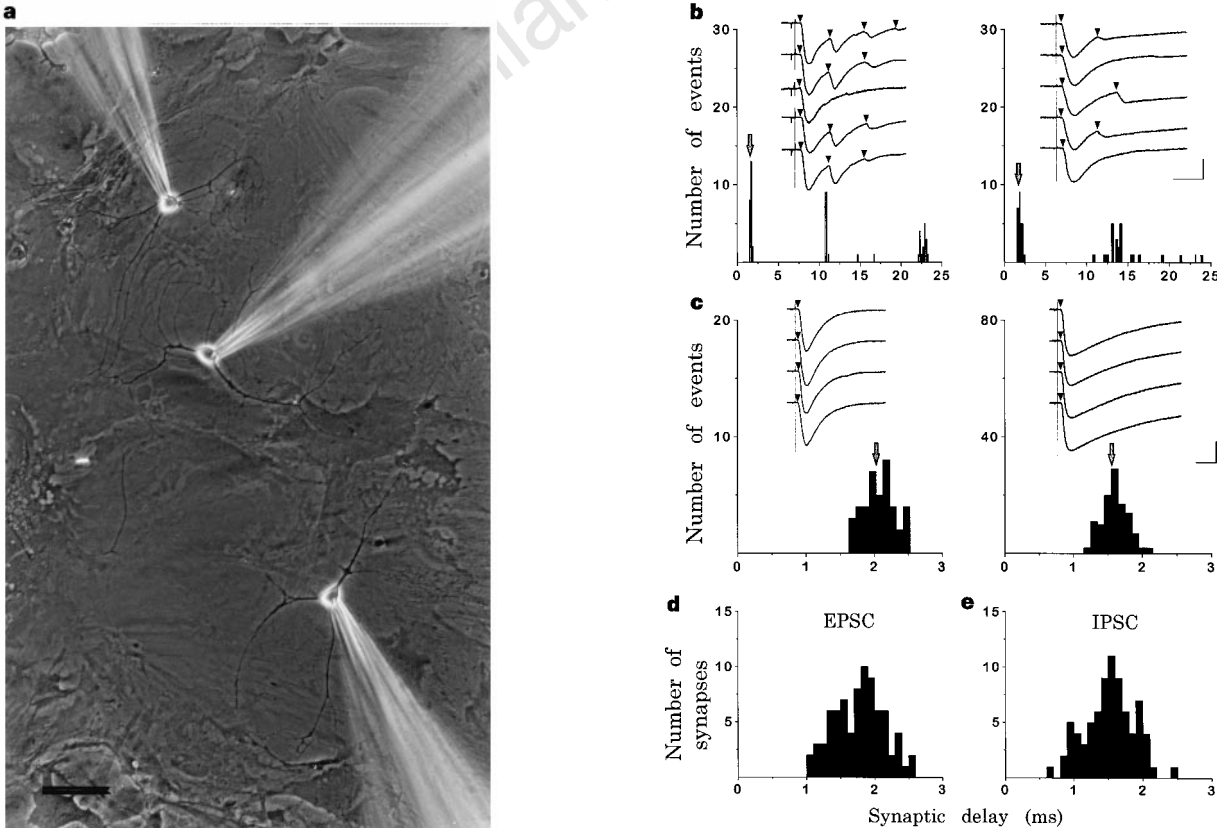


Figure 1 Simple networks of hippocampal neurons in culture. **a**, A phase-contrast microscopic image of a triplet network in a 10-day-old hippocampal culture, together with three pipettes used for whole-cell perforated patch recording. Scale bar, 50 μm . **b**, **c**, Distribution of the latency of onset of EPSCs and IPSCs recorded from four different pairs of neurons. Whole-cell recordings were made at the soma of both pre- and postsynaptic neurons in voltage-clamp mode ($V_c = -80\text{ mV}$). Step-depolarizations (+100 mV, 1 ms) were applied to the presynaptic neuron and postsynaptic membrane currents were monitored. Samples of successive

evoked synaptic currents are shown in the inset associated with each histogram. The latency was defined as the time interval between the end of 1-ms stimuli and the onset of synaptic currents. Arrowheads mark the onset time for mono- or polysynaptic responses. In **c**, arrows mark the mean latency for monosynaptic EPSCs (left) and IPSCs (right). Note the occasional failures in evoked polysynaptic responses in **b**. Scales: 20 ms, 200 pA for **b**; 10 ms, 200 pA for **c**. **d**, **e**, Distribution of the mean latency for a large number of monosynaptic glutamatergic (**d**) and GABAergic (**e**) connections ($n = 80$ each) within the triplets used here.

hippocampal neurons were simultaneously recorded and the pre-synaptic neuron was stimulated repetitively under current-clamp to fire 1-s trains of action potentials (5 Hz) with 1-s intervals for 6 min (a total of 900 stimuli) in the presence of a sustained postsynaptic depolarization ($V_c = -50$ mV). This paired stimulation gave a persistent reduction in the amplitude of the EPSCs or IPSCs at glutamatergic and GABAergic synapses, respectively (Fig. 2a, b). The reduction during the first 10–15 min following the stimulation was $49.5 \pm 0.4\%$ (s.e.m., $n = 10$) and $34.6 \pm 0.5\%$ (s.e.m., $n = 6$) for EPSCs and IPSCs, respectively, and it persisted for at least 20 min (Fig. 2c, d). Concurrent postsynaptic depolarization was necessary for the induction of LTD by the present stimulation protocol; when the postsynaptic neurons were voltage-clamped at -80 mV during the period of repetitive presynaptic stimulation, no LTD was observed (Fig. 2c, d). Furthermore, the induction of LTD at glutamatergic synapses was largely abolished by the presence of ($\pm$)-2-amino-5-phosphonopentanoic acid (AP5, $50 \mu\text{M}$; Fig. 2e), an antagonist that blocks the NMDA (*N*-methyl-D-aspartate) sub-type of glutamate receptors, whereas LTD at GABAergic synapses was unaffected by the same AP5 treatment (Fig. 2f). These properties of the induction of LTD agree with those previously described for glutamatergic synapses in hippocampal cultures²⁶ and

slices^{28–32}. The properties of GABAergic LTD remain to be further characterized.

Back propagation of synaptic depression

Serially connected triplets were first used for studying the propagation of LTD. Figure 3a shows an example consisting of only glutamatergic synapses. Matrix tabulation of representative averages of synaptic and autaptic currents is used to depict the connectivity and time-dependent synaptic changes. Alternating test stimuli were applied to neurons 1 and 2 at low frequency (0.06 Hz) to monitor the synaptic efficacy of connections, except during the 6-min period when LTD was induced at the glutamatergic synapse (E) from neuron 2 to 3 (E_{2-3}). During the induction period, neuron 2 was stimulated repetitively to fire action potentials (see above), while neuron 3 was voltage-clamped at -50 mV. Neuron 1 remained voltage-clamped at -80 mV throughout the experiment. Immediately after the stimulation, the amplitude of EPSCs at E_{2-3} was persistently reduced. EPSCs at E_{1-2} were also reduced, indicating a back propagation of depression from the axonal output of neuron 2 to its dendritic input. In a different triplet consisting of both glutamatergic and GABAergic neurons (Fig. 3b), induction of LTD at E_{2-3} was also accompanied by a depression of the GABAergic

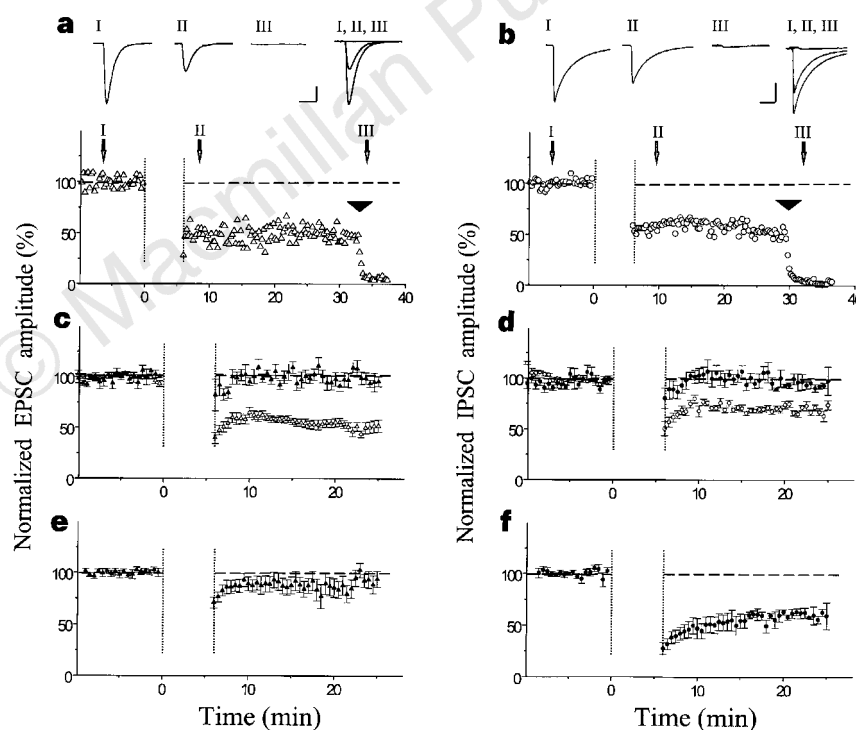


Figure 2 Induction of LTD at glutamatergic and GABAergic synapses in hippocampal cultures. **a, b**, Examples of recording at glutamatergic (**a**) and GABAergic (**b**) synapses. Data points depict the amplitudes of EPSCs or IPSCs before and after the induction of LTD, normalized to the mean value recorded (as 100%) before the induction. For the induction of LTD, the presynaptic neuron was stimulated repetitively under current clamp to fire 1-s trains of action potentials (5 Hz) with 1-s intervals for 6 min (a total of 900 stimuli) in the presence of a sustained postsynaptic depolarization ($V_c = -50$ mV). Dotted lines delineate the period of repetitive stimulation. Arrowheads indicate the time of application of glutamate antagonists ($50 \mu\text{M}$ AP5; $10 \mu\text{M}$ CNQX) and GABA_A receptor antagonists ($10 \mu\text{M}$ bicuculline) in **a** and **b**, respectively. Traces of EPSCs and IPSCs (average of 5 consecutive events) before (I) and after (II) the induction of LTD and after addition of receptor antagonists (III) are shown above. Note the slower decay of the IPSCs. Scales: 20 ms, 200 pA for **a**; 40 ms, 200 pA for **b, c, d**, Summary of

experiments on the induction of LTD at glutamatergic (**c**) and GABAergic synapses (**d**). When the postsynaptic cell was voltage-clamped at -50 mV during the repetitive presynaptic stimulation, there was significant and persistent reduction of EPSCs (**c**, open triangles; $n = 10$) and IPSCs (**d**, open circles; $n = 6$) ($P < 0.05$, Student's paired *t*-test). In contrast, when the postsynaptic neuron was voltage-clamped at -80 mV during stimulation, no significant change in EPSCs (**c**, filled triangles; $n = 5$) or IPSCs (**d**, filled circles; $n = 5$) was observed ($P > 0.05$, Student's paired *t*-test). Data points represent mean $\pm$ s.e.m. **e, f**, Effect of AP5 on the induction of LTD. The same as that described in **c** and **d**, except that AP5 ($50 \mu\text{M}$) was added in the extracellular bath solution throughout the experiment. The degree of depression was greatly reduced for glutamatergic LTD (**e**; $n = 6$), whereas GABAergic LTD was unaffected (**f**; $n = 4$). Both the residual depression at glutamatergic synapses and the depression at GABAergic synapses were significant ($P < 0.05$, Student's paired *t*-test).

synapse (I) from neuron 1 to 2 (I_{1-2}). Note that there was no depression of the autapse made by neuron 1 (E_{1-1} or I_{1-1} ; Fig. 1a, b), but the reciprocal synapse E_{2-1} was depressed in both cases. The depression of E_{2-1} suggests a lateral propagation of LTD to divergent axonal outputs of the presynaptic neuron (see below).

The results from all serial triplets recorded are summarized in Fig. 3c, d and Table 1. In control experiments (Fig. 3c), we found no change in the synaptic efficacy of either E_{1-2} or I_{1-2} after applying repetitive stimulation to neuron 2 while voltage-clamping neuron 3 at -80 mV. Thus coincident pre- and postsynaptic excitation is required not only for the induction of LTD, but also for the back propagation of depression. The same stimuli resulted in a persistent depression of both synapses E_{2-3} and E_{1-2} (or I_{1-2}) when neuron 3 was concurrently depolarized to -50 mV (Fig. 3d). The average steady depression at back-propagation site E_{1-2} or I_{1-2} was significantly lower than that of the LTD induced at E_{2-3} ($P < 0.02$, Student's t -test; Fig. 3d and Table 1). In these experiments, we noted in many cases an initial short-term depression following the

termination of the repetitive stimuli at the site of LTD induction but not at the propagation site (Fig. 3b), consistent with the absence of repetitive presynaptic stimulation at the propagation site.

In contrast to the back propagation of depression associated with glutamatergic LTD, we found no evidence for back propagation of depression following induction of LTD at GABAergic synapses. In the example shown in Fig. 4a, when LTD was induced at the GABAergic synapse I_{2-3} by the same stimulation protocol, no depression of E_{1-2} was observed. Back propagation was also absent in serial triplets that consisted only of GABAergic synapses. Neither was I_{2-1} affected by the induction of LTD at I_{2-3} , suggesting the absence of presynaptic lateral spread of depression in GABAergic neurons (see below). The results on the back propagation of GABAergic LTD are summarized in Fig. 4b and Table 1.

Forward propagation

Serial triplets of hippocampal neurons were also used to determine

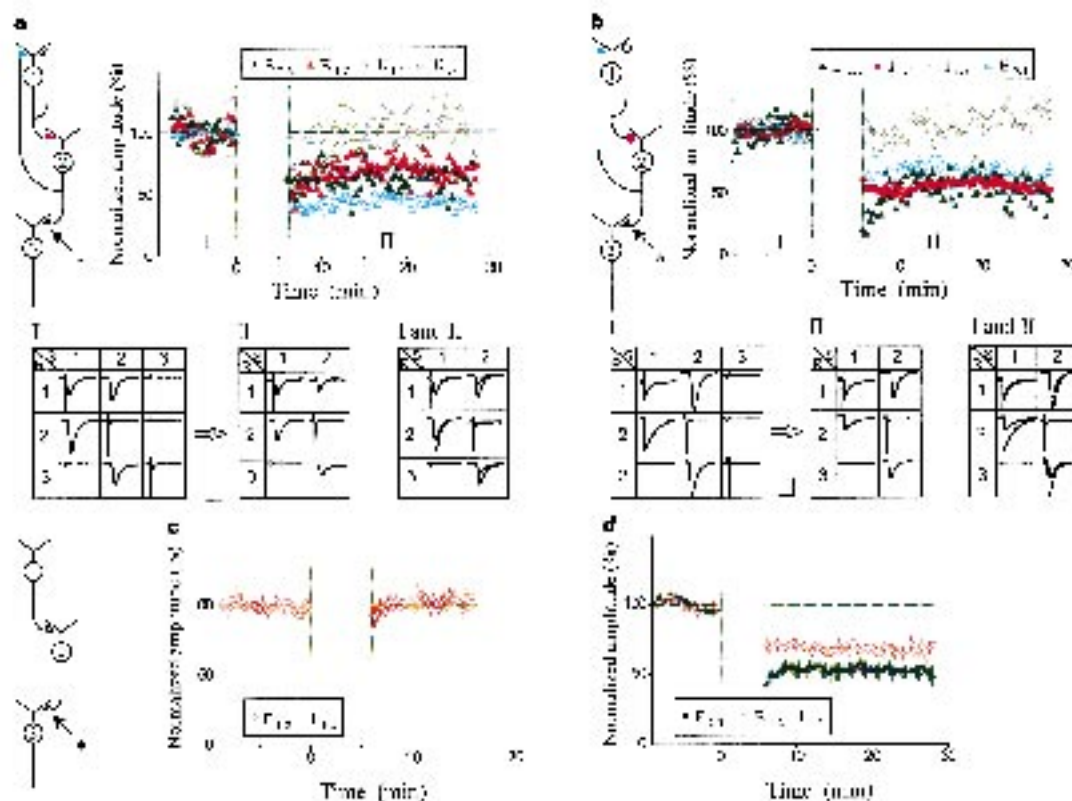


Figure 3 Back propagation of synaptic depression following the induction of LTD at glutamatergic synapses. **a**, An example of a glutamatergic triplet. The diagram on the left indicates the connectivity deduced from measurements of mono-synaptic and autaptic currents recorded from all three neurons. The matrices below depict samples (average of 4 consecutive events) of synaptic and autaptic currents before (I) and after (II) the induction of LTD as well as superimposed currents (I and II) for comparison. S, stimulated neuron; R, recorded neuron. Note that a large stimulation artefact precedes the autaptic current. Scales: 20 ms, 250 pA. Data points depict the amplitude of EPSCs at the autapse of neuron 1 (E_{1-1} , white symbols) and at synapses from neuron 1 to 2 (E_{1-2} , red), from 2 to 3 (E_{2-3} , black) and from 2 to 1 (E_{2-1} , blue). Amplitudes were normalized as for Fig. 2. The same stimulation procedure as for Fig. 2 was used to induce LTD at E_{2-3} (black arrow and asterisk), while neuron 1 was voltage-clamped at -80 mV. **b**, An example of a triplet formed by glutamatergic and GABAergic neurons. The same

as in **a**, except that neuron 1 was GABAergic. Note the slow time course for IPSCs elicited by neuron 1 (S1; scales: 40 ms, 200 pA) compared to that of EPSCs elicited by neuron 2 (S2; scales: 20 ms, 200 pA for S2 and S3). **c**, Summary of control experiments. The same repetitive stimulation as that used in the induction of LTD (**a**) was applied to neuron 2 in serial triplets while neuron 3 was voltage-clamped at -80 mV. Left, generic configuration for serial triplets studied. Data (mean $\pm$ s.e.m.) represent synaptic responses at E_{1-2} and I_{1-2} ($n = 7$, including 3 E_{1-2} and 4 I_{1-2}). **d**, Summary of back propagation of depression following the induction of LTD at glutamatergic synapses in serial triplets. The same repetitive stimulation as that used in **a** was applied to neuron 2 while neuron 3 was depolarized to -50 mV. Data points are mean $\pm$ s.e.m. ($n = 8$). Filled triangles, EPSCs at E_{2-3} ; red diamonds, averages of EPSCs and IPSCs at E_{1-2} ($n = 6$) and I_{1-2} ($n = 2$), respectively.

Figure 4 Absence of back propagation of depression following the induction of LTD at GABAergic synapses. **a**, An example of a triplet consisting of both glutamatergic and GABAergic neurons. LTD was induced at the GABAergic synapse I_{2-3} (black) by using the same protocol as for Fig. 3, and the synaptic efficacy at E_{1-2} (red) and I_{2-1} (blue) were monitored before and after induction. Scales: 40 ms, 250 pA for S2; 20 ms, 250 pA for S1 and S3. **b**, Summary of back propagation of depression following the induction of LTD at GABAergic synapses (I_{2-3}) in serial triplets similar to that in **a**. Data represent mean $\pm$ s.e.m. ($n = 7$). Black circles, IPSCs at I_{2-3} ; red diamonds, average synaptic currents at E_{1-2} ($n = 4$) and I_{2-1} ($n = 3$).

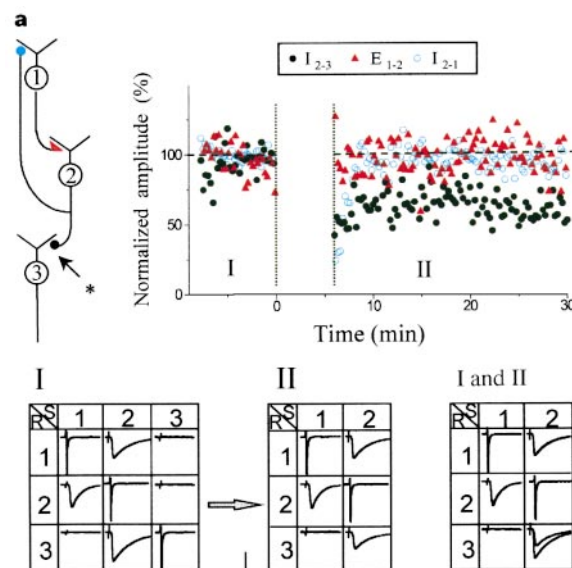
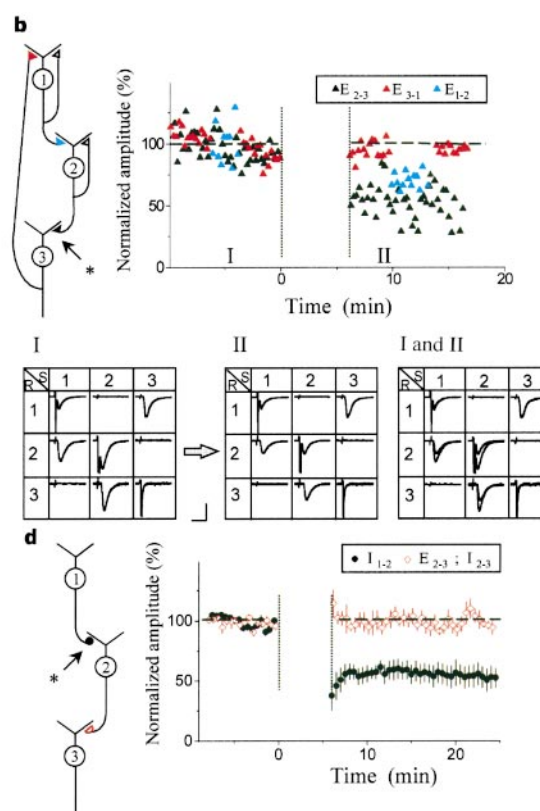
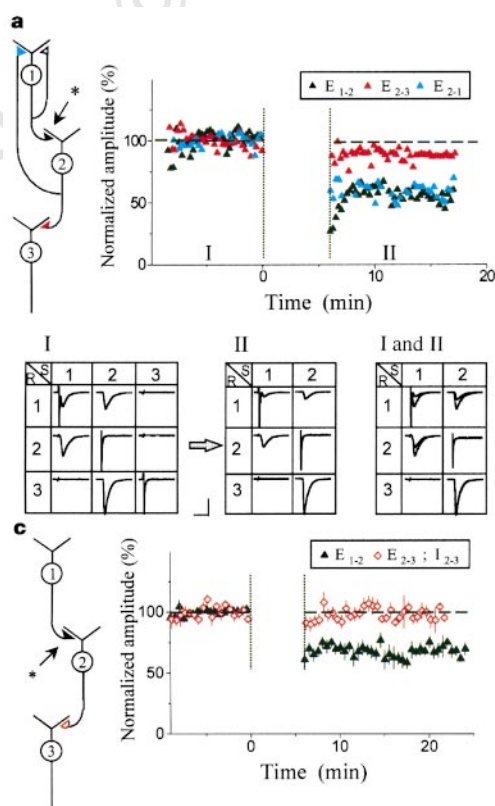
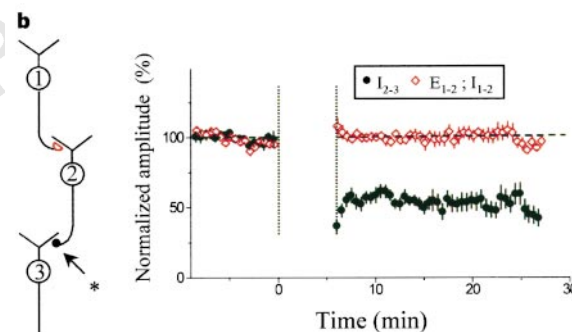


Figure 5 Absence of forward propagation of depression. **a**, An example of a serially connected triplet. Using the same stimulation procedure as that described in Fig. 3, LTD was induced at E_{1-2} . The data shown are normalized amplitudes of synaptic currents at E_{1-2} (black), E_{2-3} (red), and E_{2-1} (blue). Scale: 20 ms; 500 pA for R1, 250 pA for R2, 1 nA for R3. **b**, An example of a triplet with 'circular' connection. Alternating assays of EPSCs at E_{3-1} (red) and E_{1-2} (blue) showed that, following the induction LTD at E_{2-3} (black), depression was observed at E_{1-2} , but not at E_{3-1} . Scale: 20 ms; 500 pA for R1, 125 pA for R2 and R3. **c, d**, Summary of all experiments on forward propagation. Schematic diagram on the left depicts the generic serial connectivity used in these studies. Induction of LTD was achieved at E_{1-2} (**c**) or I_{1-2} (**d**), respectively, and the efficacy of either E_{2-3} or I_{2-3} was examined (red diamonds). Data represent mean $\pm$ s.e.m. ($n = 6$ each in **c** and **d**, including 3 E_{2-3} and 3 I_{2-3}).



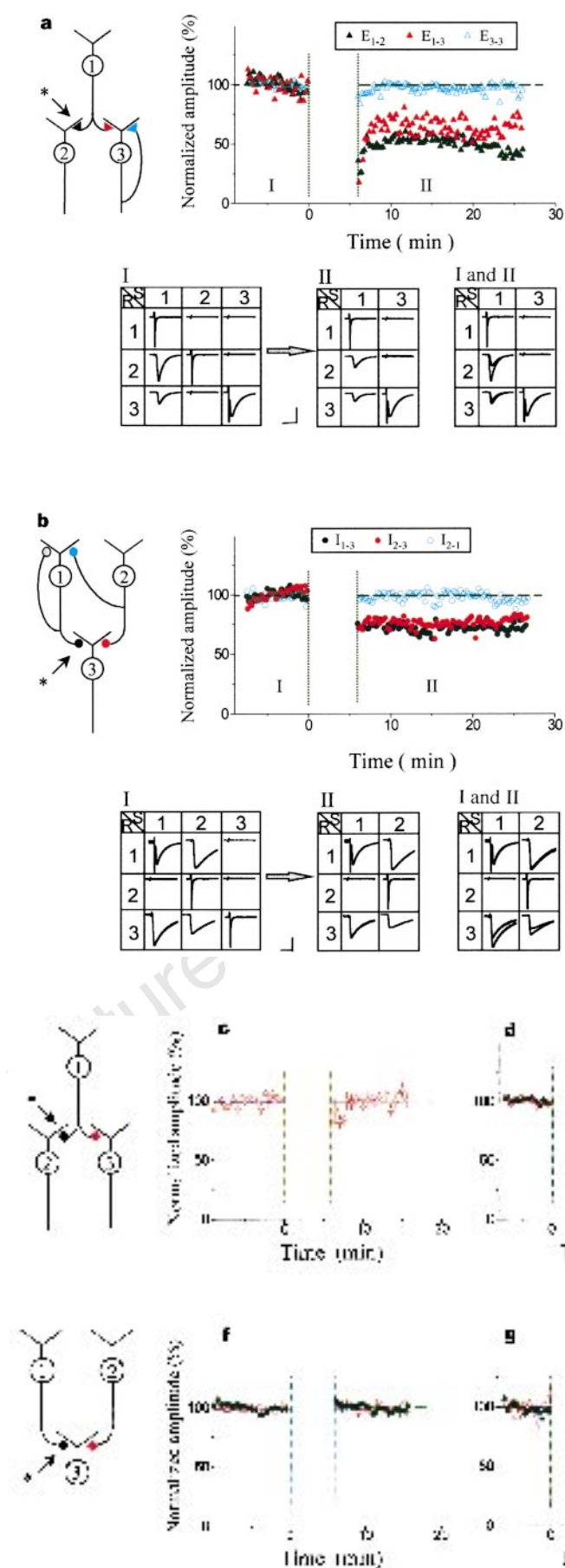


Figure 6 Lateral propagation of LTD. **a**, An example of a triplet with divergent glutamatergic outputs. For induction of LTD at E_{1-2} , neurons 1 and 2 were co-activated as for Fig. 3 while neuron 3 was voltage-clamped at -80 mV. Data shown are normalized EPSCs recorded at E_{1-2} (black), E_{1-3} (red), and E_{2-3} (blue). Scales: 20 ms; 1 nA for R1, 500 pA for R2 and R3. **b**, An example of a triplet with converging GABAergic inputs. Induction of LTD was achieved at I_{1-3} while neuron 2 was voltage-clamped at -80 mV. Data depict normalized amplitudes of EPSCs at I_{1-3} (black), I_{2-3} (red) and I_{2-1} (blue), before and after induction of LTD at I_{1-3} . Scales: 20 ms; 400 pA for R1 and R2, 100 pA for R3. **c-h**, Summary of results from experiments on lateral propagation of depression. Diagrams on the left depict the generic triplet configuration used. **c**, Control experiments. Repetitive stimulation of the presynaptic glutamatergic neuron (as in **a**) in the absence of postsynaptic depolarization of neuron 2 and 3 ($V_c = -80$ mV) did not induce synaptic depression at E_{1-3} . Data represent mean $\pm$ s.e.m. ($n = 6$). **d**, Summary of presynaptic lateral propagation of depression following induction of glutamatergic LTD. Data shown are normalized amplitudes of EPSCs (mean $\pm$ s.e.m.; $n = 6$) at the site of LTD induction (black) and at the other divergent output of the same presynaptic neuron (red). **e**, Summary of presynaptic lateral propagation of depression following the induction of GABAergic LTD. Shown are data of normalized amplitudes of IPSCs at the site of LTD induction (black) and at the other divergent output (red) ($n = 4$). **f**, Control experiments. Sustained depolarization of the postsynaptic cell ($V_c = -50$ mV) for 6 min in the absence of repetitive presynaptic stimulation did not induce any change in EPSCs (red, $n = 4$) or IPSCs (black, $n = 7$). Data represent mean $\pm$ s.e.m. **g**, Summary of postsynaptic lateral propagation of depression following the induction of glutamatergic LTD. Data depict normalized amplitudes of EPSCs at the site of LTD induction (black) and of EPSCs ($n = 5$) or IPSCs ($n = 1$) at the other convergent input on the same postsynaptic neuron (red). Data represent mean $\pm$ s.e.m. ($n = 6$). **h**, Summary of postsynaptic lateral propagation following the induction of GABAergic LTD. Data depict normalized amplitudes of IPSCs at the site of LTD induction (black) and of EPSCs ($n = 3$) or IPSCs ($n = 1$) at the other convergent input on the same postsynaptic neuron (red). Data represent mean $\pm$ s.e.m. ($n = 4$).

whether LTD can spread in the anterograde direction from the dendrites of the postsynaptic neuron to synapses made by its axonal terminals. In the example shown in Fig. 5a, we found that induction of LTD at E_{1-2} did not result in significant depression of E_{2-3} , although the depression had back propagated to E_{2-1} and laterally propagated to E_{1-1} (see matrix in Fig. 5a). In an example of a 'circular' triplet (Fig. 5b), induction of LTD at E_{2-3} did not result in any depression of E_{3-1} , but caused significant depression of E_{1-2} . Thus back propagation occurred in the absence of forward propagation within the same triplet. In other experiments on serial triplets containing GABAergic neurons, we also found no evidence of forward propagation of depression to either E_{2-3} or I_{2-3} following the induction of LTD at I_{1-2} . These results on the absence of forward propagation are summarized in Fig. 5c, d and Table 1.

Lateral propagation

In addition to the serial configuration, networks of triplets with other configurations were also studied. In triplets consisting of one neuron making connections onto two other neurons (Table 1), we examined whether LTD can spread laterally between divergent outputs of the presynaptic neuron. An example of pure glutamatergic triplet is shown in Fig. 6a. LTD was induced at E_{1-2} by repetitive firing of neuron 1 concurrently with sustained depolarization of neuron 2 ($V_c = -50$ mV), while neuron 3 remained voltage-clamped at -80 mV. Following the induction of LTD at E_{1-2} , significant and persistent depression was also observed at E_{1-3} , while E_{3-3} was not affected. Control experiments showed that repetitive firing of the presynaptic neuron alone, with the postsynaptic neuron 2 and 3 voltage-clamped at -80 mV, resulted in no persistent change in the amplitude of EPSCs at either E_{1-2} or E_{1-3} (Fig. 6c). For all divergent triplets examined, we found consistent propagation of LTD to other synaptic outputs of the glutamatergic neuron and the degree of depression at the propagation site correlated with that found at the induction site (Figs 6d, 7e and Table 1). Triplet networks consisting of GABAergic neurons with divergent outputs were also examined. As summarized in Fig. 6e, the same paired-stimulation of neuron 1 and 2 as described above resulted in LTD of the GABAergic synapse I_{1-2} . However, only an initial transient depression of I_{1-3} was observed following the induction of LTD at I_{1-2} , indicating an absence of lateral propagation of LTD to divergent GABAergic outputs. The lack of presynaptic lateral spread was also observed for the triplet shown in Fig. 4a. Note that the appearance of transient depression at all divergent outputs of the presynaptic neuron is consistent with short-term depression following repetitive stimulation. Overall, although the extent of LTD at GABAergic synapses was similar to that found at glutamatergic synapses, there was no spread of depression to divergent outputs of the presynaptic GABAergic neurons (Table 1).

We have also examined triplets that contained two neurons making convergent inputs on a third neuron (Table 1). An example of a triplet with two convergent GABAergic inputs is shown in Fig. 6b. Induction of LTD at I_{1-3} was achieved by repetitive firing of neuron 1 in the presence of sustained depolarization of neuron 3 at -50 mV, while neuron 2 was voltage-clamped at -80 mV. This also resulted in a persistent depression of I_{2-3} , indicating lateral propagation of LTD to convergent inputs on neuron 3. However, the autapse I_{1-1} and synapse I_{2-1} were not affected by the induction of LTD at I_{1-3} , consistent with the lack of presynaptic lateral and back propagation in the GABAergic neurons. Sustained postsynaptic depolarization of neuron 3 during the induction of LTD was not sufficient to cause the depression at I_{2-3} , because depolarization of postsynaptic cells at -50 mV for 6 min by itself had no effect on either glutamatergic or GABAergic synapses (Fig. 6f). Figure 6g, h summarizes results of all experiments on postsynaptic lateral propagation following the induction of LTD at glutamatergic and GABAergic synapses, respectively. Depression was similar for

lateral propagation to either glutamatergic or GABAergic synapse (Table 1).

Range of propagation

Our simultaneous recordings were limited to only three neurons; thus the propagation of depression was not monitored beyond the triplet network. We noted, however, that the autapse E_{1-1} in triplets shown in Figs 3a, b and 5b was not depressed, whereas E_{1-2} were depressed by back propagation in presynaptic neuron 2. Assuming that autapses may be considered as converging inputs or divergent output within the network, then the absence of depression at E_{1-1} indicates that propagated depression at E_{1-2} does not further laterally or back propagate to a more 'upstream' neuron 1. In the divergent triplet shown in Fig. 6a, the autapse E_{3-3} did not become depressed while depression had propagated to E_{1-3} . Further lateral propagation was lacking in all cases with similar configuration ($n = 5$). The propagated depression at the divergent output thus appears to be incapable of further propagation to more 'downstream' synapses made by or on the postsynaptic neurons associated with that divergent output. In addition, we found that in all cases examined for postsynaptic lateral propagation ($n = 7$), autapses in neurons not directly associated with the induction of LTD were not depressed, regardless of whether LTD was induced at glutamatergic or GABAergic synapses. The lack of further propagation beyond the immediate pre- and postsynaptic neurons involved in the induction of LTD may be due to the limited degree of propagated depression. Alternatively, there may be a qualitative difference in the mechanism underlying the depression at the induction and propagation sites.

Potential cellular mechanisms

We have observed extensive propagation of depression from the site of induction of LTD to other synaptic sites within a triplet network. The propagation may result from the spread of 'depressive' signals either within the pre- or postsynaptic neurons or through the extracellular space. In these cultures, multiple synaptic (and autaptic) contacts made on each neuron were usually distributed rather haphazardly. The average spatial distributions of synaptic and autaptic sites with respect to sites of LTD induction are likely to be similar, regardless of their functional position within the serial network. However, the depression did not spread non-selectively to all synapses and autapses within the network. This was shown by the absence of forward propagation, whereas lateral and back propagation were observed within the same triplet (Fig. 5) and by the lack of effect on autapses formed on the neuron exhibiting propagated depression in the example shown in Figs 3, 5b, 6a. These findings, together with the fact that constant perfusion of fresh medium was applied to the culture throughout the recording period, are evidence against extracellular diffusible factors being mediators of the spread of depression.

Previous studies of LTD at developing neuromuscular junctions^{33,34} and in brain slices^{29,30,35} have shown that, although postsynaptic influx of Ca^{2+} is required for its induction, LTD is expressed as a presynaptic reduction of evoked transmitter release. This suggests the existence of retrograde signalling from the post- to the presynaptic neuron. The retrograde signal may generate a secondary cytoplasmic factor in the presynaptic glutamatergic neuron for the back and lateral propagation of depression. The absence of back and lateral propagation of GABAergic LTD suggests a postsynaptic expression of LTD at these synapses, without involvement of retrograde presynaptic modification. As there is no forward propagation of depression for both glutamatergic and GABAergic synapses, the depressive signal produced at the site of LTD induction appears to be incapable of propagation in the anterograde direction. The postsynaptic site may fail to generate a necessary secondary signal, the signal may not be transported in the anterograde direction, or the output terminal of the postsynaptic cell may not be susceptible to modification by the signal.

Finally, although cytoplasmic signalling would explain a selective propagation of depression, we cannot exclude the possibility that extracellular depressive signals are involved. Repetitive paired stimulation during the induction of LTD may have rendered all synapses associated with the pre- or postsynaptic neuron selectively susceptible to an extracellular depressive signal.

A clue to the mechanisms underlying back propagation of depression may be provided by the extent of back-propagated depression in different triplets. As shown in Fig. 7a, the degree of depression observed at either E_{1-2} or I_{1-2} after the induction of LTD at E_{2-3} was strongly correlated with the initial amplitude of EPSCs or IPSCs at E_{1-2} or I_{1-2} , respectively. A weak correlation was found between the degree of the propagated depression and the rise time of the synaptic currents (Fig. 7b). In contrast, for either the induced LTD or forward-propagated depression, there was no correlation between the degree of depression and the initial amplitude of the synaptic current (Fig. 7c). The amplitude and rise time may reflect the average electrotonic distance of various synaptic sites from the soma; those on more distal dendrites elicit currents of smaller amplitude and longer rise time. The correlation found for the back-propagated depression is thus consistent with the notion of cytoplasmic retrograde signals, which decrement with the cytoplasmic distance between the axonal endings and synaptic sites on the dendrite. On the other hand, the amplitude of the synaptic current can also be attributed to other intrinsic properties of the synapse, for example the number of synaptic boutons, the effectiveness of the

synaptic contact, or the probability of transmitter secretion. These properties may result in a differential susceptibility of a synapse to the retrograde signal, resulting in the observed correlation between the degree of depression at the back-propagation site and the initial amplitude of the postsynaptic currents at that site. Indeed, synapses with different release probability show different degrees of short-term plasticity³⁶.

The presence of reciprocal connections between neurons 1 and 2 (E_{1-2} and E_{2-1}) was neither required nor related to the degree of propagated depression, as shown by the similar range of depression for triplets without E_{2-1} (Figs 7a, b, 5b). However, we cannot exclude the potential contribution of recurrent presynaptic inhibition or facilitation, which escapes detection by our recording methods. The mean amplitude of synaptic currents for synapses used for examining forward propagation covered a range of values similar to that used for studying back propagation (Fig. 7c), indicating that the lack of forward propagation was not due to a biased sampling of forward synapses with smaller synaptic currents. We found no correlation between the degree of back-propagated depression and induced depression (Fig. 7d), suggesting that the extent of induced LTD bears no relation to the retrograde influence received by the dendrite of the presynaptic neuron. In contrast, the extent of depression due to presynaptic lateral propagation is linearly correlated with that of the induced LTD (Fig. 7e). There was no obvious correlation between the postsynaptic lateral depression and the induced depression (Fig. 7f). These differences indicate

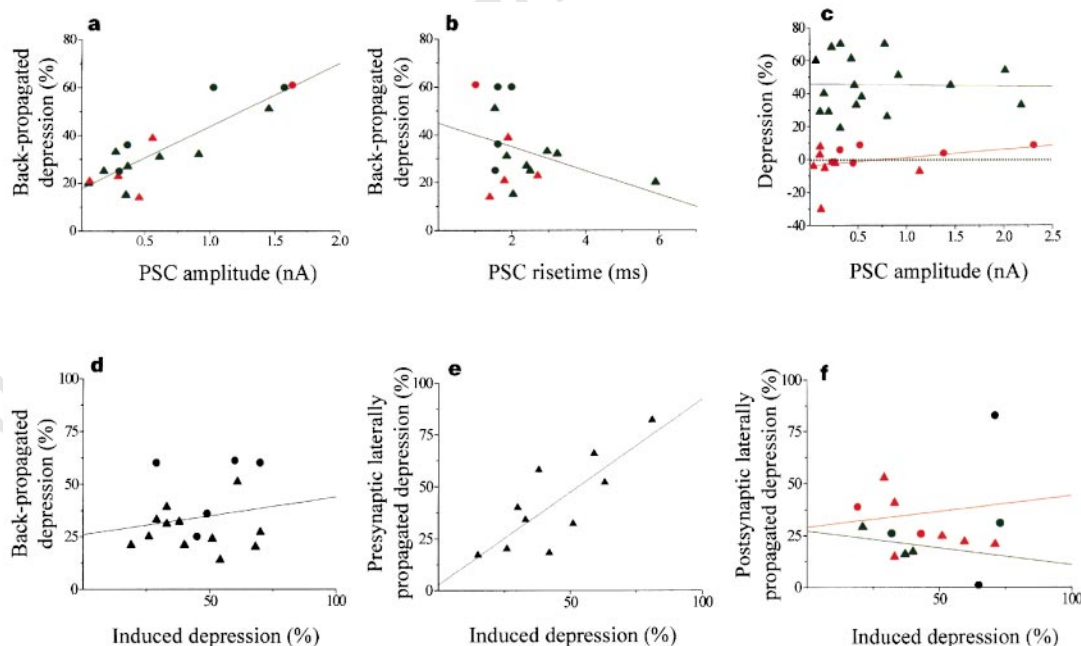


Figure 7 **a, b**, Correlation between the degree of back propagation of synaptic depression and synaptic properties. The percentage of reduction in the mean amplitude of EPSCs (triangles, $n = 12$) or IPSCs (circles, $n = 5$) at the back-propagation site of presynaptic neurons during the first 10–15 min following the induction of LTD is plotted against the initial mean amplitude (**a**) or the risetime (**b**) of the synaptic currents at the back-propagation sites before induction. The same data set as that shown in Table 1 was used. The line represents the best linear fit of the data (correlation coefficient, $r = 0.87$, $P < 0.001$, and $r = -0.36$, $P = 0.16$ in **a** and **b**, respectively). Red symbols represent serial triplets that did not contain the reciprocal connection E_{2-1} (Fig. 3). **c**, The degree of depression for the induced LTD (black) or forward-propagated depression (red) is plotted against the initial amplitude of the synaptic current at the site of induction or the forward-propagation site, respectively. Triangles and circles represent glutamatergic and GABAergic synapses, respectively. Lines represent the best linear fits to the data (black,

$r = -0.02$, $P = 0.9$; red, $r = 0.32$, $P = 0.3$). **d**, The degree of depression at the back-propagated site is plotted against that observed at the site of LTD induction for each triplet. Triangles and circles represent glutamatergic and GABAergic synapses, respectively. Lines represent best linear fits to the data ($r = 0.20$, $P = 0.5$). **e**, The degree of depression at the presynaptic propagated sites and LTD induction sites was plotted for each triplet network examined for glutamatergic synapses. The line represents the best fit to the data ($r = 0.80$, $P = 0.005$). **f**, The degree of depression at the propagation and induction sites was plotted for postsynaptic lateral propagation of LTD to glutamatergic (triangles) and GABAergic (circles) synapses. Red and black symbols represent data for which LTD was induced at a glutamatergic or GABAergic synapse, respectively, and the corresponding lines represent the best linear fits to the data (red, $r = 0.13$, $P = 0.7$; black, $r = -0.28$, $P = 0.6$).

that the nature of cellular signals underlying various forms of pre- and postsynaptic propagation of depression is likely to be different.

Information flow in a neural network

Information coded by changes in the membrane potential flows rapidly from dendrites to the axon and across the synapse to the postsynaptic neuron in an anterograde direction. Retrograde signals also flow across the synapse on a slower timescale to regulate the development, maintenance and modification of the properties of the presynaptic neuron^{1,37–41}. Gradual loss of synapses at the dendrites after axotomy^{42–44} or interruption of axonal transport⁴⁵ has demonstrated the existence of slow axon-dendritic trophic interaction of the order of days to weeks which is crucial for maintaining the structural and functional integrity of synaptic inputs at the dendrites. Our results on the back propagation of synaptic depression further indicate an axon-dendritic signalling that conveys information on activity-dependent synaptic modification at the axonal terminals with a timescale of the order of minutes. In addition to their role in integrating synaptic potentials, dendrites may also integrate retrograde signals on the efficacy of the axonal outputs of a neuron. Further investigation is necessary to determine whether synaptic potentiation also undergoes back propagation, and whether back-propagated influences from multiple axonal outputs of the neuron can be integrated at its dendrites.

The extensive propagation of synaptic modification in glutamatergic neurons suggests that local excitatory activity within the neural network may result in a more extensive distribution of long-term synaptic modification in the network than that induced by repetitive activities at GABAergic synapses. Lateral propagation of depression, either pre- or postsynaptic, will result in a loss of pathway specificity associated with LTD. This loss of specificity appears to contradict findings on LTD induced at CA1 regions of hippocampal slices, where only synapses made by the activated input were depressed, but non-activated inputs remained unaffected^{7,28,29}. However, these latter experiments were usually done using well-separated inputs to the CA1 region. Lateral propagation of LTD may be spatially restricted, in a way similar to that found for the lateral spread of LTP in hippocampal slice^{9,10} and slice cultures^{8,11}. The extensive propagation of depression observed here may reflect the restricted geometry of nerve processes and higher density of connectivity between neurons in an artificial culture condition. Although the phenomena in cell culture may be an exaggeration of those occurring *in vivo*, our simple neural networks will enable investigation of network properties that may be present to varying degrees in more complex neural systems.

Artificial neural networks with multiple layers of interconnected 'neuronal' units and modifiable synaptic connections have been used for modelling learning processes of the nervous system²³. Such a network can be trained by reiterative adjustments of the weight of synaptic connections between units of different layers to reduce the errors in the output. An efficient computational procedure for supervised learning in simple feedforward networks, known as the back-propagation algorithm^{22,46}, uses the error in the output to compute adjustments to the weights of synaptic connections from upstream units within the network. The biological plausibility of such an algorithm requires information flow from output synapses of a neuron back towards its input synapses. Although more complex feedback circuits may be used to convey information in a retrograde direction⁴⁷, the back propagation of synaptic modification demonstrated here offers an efficient cellular mechanism for the implementation of a powerful learning algorithm in a simple biological neural network. □

Methods

Cell culture. Low-density cultures of dissociated embryonic rat hippocampal neurons were prepared as described²⁴. Briefly, hippocampi were removed from E18–20 embryonic rats and treated with trypsin for 20 min at 37 °C, followed

by washing and gentle trituration. The dissociated cells were plated at densities of 30,000 to 50,000 cells ml⁻¹ on poly-L-lysine-coated glass coverslips in 35-mm Petri dishes. The plating medium was Dulbecco's minimum essential medium (DMEM; BioWhittaker) supplemented with 10% calf serum (Hyclone), 10% Ham's F12 with glutamine (BioWhittaker) and 50 U ml⁻¹ penicillin–streptomycin (Sigma). The culture medium was changed to the above medium containing 20 mM KCl 24 h after plating. Both glial and neuronal cell types are present under these culture conditions. Cells were used for electrophysiological recordings after 10–14 days in culture.

Electrophysiology. Whole-cell perforated patch recordings^{48–50} from three hippocampal neurons were performed simultaneously, using amphotericin B (Sigma) for perforation. The micropipettes were made from borosilicate glass capillaries (Kimax), with a resistance in the range of 2–4 MΩ. The pipettes were tip-filled with internal solution and then back-filled with internal solution containing 150 ng ml⁻¹ amphotericin B. The internal solution contained the following (in mM): potassium gluconate 136.5; KCl 17.5; NaCl 9; MgCl₂ 1; HEPES 10; EGTA 0.2 (pH 7.30). The external bath solution was a HEPES-buffered saline (HBS) containing the following (in mM): NaCl 145; KCl 3; HEPES 10; CaCl₂ 3; glucose 8; MgCl₂ 2 (pH 7.30). The bath was constantly perfused with fresh recording medium at a slow rate throughout the recording; all experiments were done at room temperature. Neurons were visualized by phase-contrast microscopy with a Nikon inverted microscope. Recordings were made with three patch-clamp amplifiers (Axopatch 200; Axon Instruments). Signals were filtered at 5 kHz using amplifier circuitry and stored on a VCR. Data were sampled at 10 kHz and analysed using a pClamp 6.0 software (Axon Instruments). Series resistance (10–30 MΩ) was always compensated at 75–80% (lag, 10 μs). Data from experiments were accepted for analysis only in cases where the postsynaptic currents did not vary beyond 10% of the average values during the control period and the input resistance (100–300 MΩ) remained constant throughout the experiment. For assaying synaptic connectivity, each neuron was stimulated at a low frequency (0.06 Hz) by 1-ms step depolarization from –80 mV to +20 mV in voltage-clamp mode, and the responses from the other two neurons as well as autaptic responses in the stimulated neuron itself were recorded. The use of potassium gluconate as the principal ionic constituent in the intracellular solution resulted in inward EPSCs and IPSCs at resting membrane potentials (–80 to –60 mV). The IPSCs have distinctly longer decay times and more negative reversal potentials than EPSCs. Autaptic currents, when present, followed the 1-ms step depolarization in voltage-clamp mode. Consistent with previous reports^{24,26}, EPSCs and IPSCs in these cultures are glutamatergic and GABAergic in nature, respectively. As shown in Fig. 2a, b, pharmacological studies indicated that EPSCs are mediated by AMPA and NMDA receptors, blocked by receptor antagonists (±)-2-amino-5-phosphonopentanoic acid (AP5, 50 μM; Research Biochemicals International) and 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, 10 μM, Research Biochemicals International); IPSCs are mediated by GABA_A receptors, blocked by bicuculline methiodide (10 μM, Research Biochemicals International). For all triplets consisting of divergent outputs (configuration 2 in Table 1), in no case did we observe a single neuron making both glutamatergic and GABAergic synapses with other neurons, consistent with the expected uniformity in the transmitter type used at all axonal terminals of the same neuron.

Received 31 December 1996; accepted 11 June 1997.

1. Purves, D. & Lichtman, J. W. *Principles of Neural Development* (Sinauer Associates, Sunderland, MA, 1985).
2. Goodman, C. & Shatz, C. J. Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell/Neuron* **72**, 77–98 (1993).
3. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
4. Bliss, T. V. & Lomo, T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol. (Lond.)* **232**, 331–356 (1973).
5. Ito, M. The cellular basis of cerebellar plasticity. *Curr. Opin. Neurobiol.* **1**, 616–620 (1991).
6. Bliss, T. V. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
7. Linden, D. J. & Connor, J. A. Long-term synaptic depression. *Annu. Rev. Neurosci.* **18**, 319–357 (1995).
8. Bonhoeffer, T., Staiger, V. & Aertsen, A. Synaptic plasticity in rat hippocampal slice cultures: local 'Hebbian' conjugation of pre- and postsynaptic stimulation leads to distributed synaptic enhancement. *Proc. Natl Acad. Sci. USA* **86**, 8113–8117 (1989).
9. Schuman, E. M. & Madison, D. V. Locally distributed synaptic potentiation in the hippocampus. *Science* **263**, 532–536 (1994).
10. Kossel, A., Bonhoeffer, T. & Bolz, J. Non-Hebbian synapses in rat visual cortex. *Neuroreport* **1**, 115–118 (1990).

11. Engert, F. & Bonhoeffer, T. Synapse specificity of long-term potentiation breaks down at short distances. *Nature* **388**, 279–284 (1997).
12. Scanziani, M., Malenka, R. C. & Nicoll, R. A. Role of intercellular interactions in heterosynaptic long-term depression. *Nature* **380**, 446–450 (1996).
13. McMahon, L. L. & Kauer, J. A. Hippocampal interneurons express a novel form of synaptic plasticity. *Neuron* **18**, 266, 737–739 (1997).
14. Lynch, G. S., Dunwiddie, T. & Gribkoff, V. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature* **266**, 737–739 (1977).
15. Abraham, W. C. & Goddard, G. V. Asymmetric relationships between homosynaptic long-term potentiation and heterosynaptic long-term depression. *Nature* **305**, 717–719 (1983).
16. Christie, B. R. & Abraham, W. C. NMDA-dependent heterosynaptic long-term depression in the dentate gyrus of anaesthetized rats. *Synapse* **10**, 1–6 (1992).
17. Otani, S., Connor, J. A. & Levy, W. B. Long-term potentiation and evidence for novel synaptic potentiation in CA1 stratum-oriens of rat hippocampus. *Learn. Mem.* **2**, 101–106 (1995).
18. Muller, D., Hefft, S. & Figurov, A. Heterosynaptic interactions between LTP and LTD in cal hippocampal slices. *Neuron* **14**, 599–605 (1995).
19. Lo, Y. & Poo, M.-m. Activity-dependent synaptic competition *in vitro*: Heterosynaptic suppression of developing synapses. *Science* **254**, 1019–1022 (1991).
20. Cash, S., Zucker, R. S. & Poo, M.-m. Spread of synaptic depression by presynaptic cytosolic signaling. *Science* **272**, 998–1001 (1996).
21. Vincent, P. & Marty, A. Neighboring cerebellar Purkinje cells communicate via retrograde inhibition of common presynaptic interneurons. *Neuron* **11**, 885–893 (1993).
22. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* Vol. 1 (eds Rumelhart, D. E. & McClelland, J. L.) 316–362 (MIT Press, Cambridge, MA, 1986).
23. Churchland, P. S. & Sejnowski, T. J. *The Computational Brain* (MIT Press, Cambridge, MA, 1992).
24. Wilcox, K. S., Buchhalter, J. & Dichter, M. A. Properties of inhibitory and excitatory synapses between hippocampal neurons in very low density cultures. *Synapse* **18**, 128–151 (1994).
25. Bekkers, J. M. & Stevens, C. F. Excitatory and inhibitory autaptic currents in isolated hippocampal neurons maintained in cell culture. *Proc. Natl Acad. Sci. USA* **88**, 7834–7838 (1991).
26. Goda, Y. & Stevens, C. F. Long-term depression properties in a simple system. *Neuron* **16**, 103–111 (1996).
27. Tong, G., Malenka, R. C. & Nicoll, R. A. Long-term potentiation in cultures of single hippocampal granule cells: a presynaptic form of plasticity. *Neuron* **16**, 1147–1157 (1996).
28. Dudek, S. M. & Bear, M. F. Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proc. Natl Acad. Sci. USA* **89**, 4363–4367 (1992).
29. Mulkey, R. M. & Malenka, R. C. Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus. *Neuron* **9**, 967–975 (1992).
30. Bolshakov, V. Y. & Siegelbaum, S. A. Postsynaptic induction and presynaptic expression of hippocampal long-term depression. *Science* **264**, 1148–1151 (1994).
31. Stevens, C. F. & Wang, Y. Changes in reliability of synaptic function as a mechanism for plasticity. *Nature* **371**, 704–707 (1994).
32. Xiao, M.-Y., Wigstrom, H. & Gustafsson, B. Long-term depression in the hippocampal CA1 region is associated with equal changes in AMPA and NMDA receptor-mediated synaptic potentials. *Eur. J. Neurosci.* **6**, 1055–1057 (1994).
33. Dan, Y. & Poo, M.-m. Hebbian depression of isolated neuromuscular synapses *in vitro*. *Science* **256**, 1570–1573 (1992).
34. Cash, S., Dan, Y., Poo, M.-m. & Zucker, R. Depression of developing neuromuscular synapses by postsynaptic elevation of calcium. *Neuron* **16**, 745–754 (1996).
35. Brocher, S., Artola, A. & Singer, W. Intracellular injection of Ca^{2+} chelators blocks induction of long-term depression in rat visual cortex. *Proc. Natl Acad. Sci. USA* **89**, 123–127 (1992).
36. Murthy, V. N., Sejnowski, T. J., Stevens, C. F. Heterogeneous release properties of visualized individual hippocampal synapses. *Neuron* **18**, 599–612 (1997).
37. Clarke, P. G. H. The role of trophic communication in biological networks. *Concepts Neurosci.* **2**, 201–219 (1991).
38. Jessell, T. M. & Kandel, E. R. Synaptic transmission: a bidirectional and self-modifiable form of cell-cell communication. *Cell/Neuron* **72/10**, 1–30 (1993).
39. Dan, Y. & Poo, M.-m. Retrograde interactions during formation and elimination of neuromuscular synapses. *Curr. Opin. Neurobiol.* **4**, 95–100 (1994).
40. Davis, G. W. & Murphey, R. K. Retrograde signaling and the development of transmitter release properties in the invertebrate nervous system. *J. Neurobiol.* **25**, 740–756 (1994).
41. Fitzsimonds, R. M. & Poo, M.-m. Retrograde signaling in the development and modulation of synapses. *Physiol. Rev.* (in the press).
42. Mathews, M. R. & Nelson, V. H. Detachment of structurally intact nerve endings from chromatolytic neurons of rat superior cervical ganglion during the depression of synaptic transmission induced by post-ganglionic axotomy. *J. Physiol. (Lond.)* **245**, 91–135 (1975).
43. Purves, D. Functional and structural changes in mammalian sympathetic neurones following interruption of their axons. *J. Physiol. (Lond.)* **252**, 429–463 (1975).
44. Wood, M. R. & Faber, D. S. Electrophysiological and morphological correlates of axotomy-induced deafferentation of the goldfish Mauthner cell. *J. Comp. Neurol.* **244**, 413–426 (1986).
45. Pilar, G. & Landmesser, L. Axotomy mimicked by localized colchicine application. *Science* **177**, 1116–1118 (1972).
46. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagation errors. *Nature* **323**, 533–536 (1986).
47. Zipser, D. & Rumelhart, D. E. in *Computational Neuroscience* (ed. Schwartz, E. L.) 192–200 (MIT Press, Cambridge, MA, 1990).
48. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
49. Horn, R. & Marty, A. Muscarinic activation of ionic currents measured by a new whole-cell recording method. *J. Gen. Physiol.* **92**, 145–159 (1988).
50. Rae, J., Cooper, K., Gates, P. & Watsky, M. Low access resistance perforated patch recordings using amphotericin B. *J. Neurosci. Meth.* **37**, 15–26 (1991).

Acknowledgements. We thank T. Mau for preparing cell cultures, and Y. Dan, B. Berninger and A. Schinder for discussion. This work was supported by grants from the NIH and the NSF.

Correspondence and requests for materials should be addressed to M.-m.P. (e-mail: mpoo@ucsd.edu).

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA. Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Quantum oscillations between two weakly coupled reservoirs of superfluid ^3He

S. V. Pereverzev*, A. Loshak, S. Backhaus, J. C. Davis & R. E. Packard

Physics Department, University of California, Berkeley, California 94720, USA

Arguments first proposed over thirty years ago, based on fundamental quantum-mechanical principles, led to the prediction^{1–3} that if macroscopic quantum systems are weakly coupled together, particle currents should oscillate between the two systems. The conditions for these quantum oscillations to occur are that the two systems must both have a well defined quantum phase, ϕ , and a different average energy per particle, μ : the term ‘weakly coupled’ means that the wavefunctions describing the systems must overlap slightly. The frequency of the resulting oscillations is then given by $f = (\mu_2 - \mu_1)/h$, where h is Planck’s constant. To date, the only observed example of this phenomenon is the oscillation of electric current between two superconductors coupled by a Josephson tunnelling weak link⁴. Here we report the observation of oscillating mass currents between two reservoirs of superfluid ^3He , the weak link being provided by an array of submicrometre apertures in a membrane separating the reservoirs. An applied pressure difference creates mass-current oscillations, which are detected as sound in a nearby microphone. The sound frequency (typically 6,000–200 Hz) is precisely proportional to the applied pressure difference, in accordance with the above equation. These superfluid quantum oscillations were first detected while monitoring an amplified microphone signal with the human ear.

The theory underlying the above equation was developed in the context of generalizing the ideas of Josephson⁴. He predicted that in a superconducting tunnel junction, the quantum phase difference, $\Delta\phi$, across the junction is related to the electrical current, I , through it, by the equation

$$I = I_c \sin(\Delta\phi) \quad (1)$$

where I_c is the critical current of the junction. The phase difference evolves in time t according to

$$\frac{\partial(\Delta\phi)}{\partial t} = -\frac{\Delta\mu}{\hbar} \quad (2)$$

where $\hbar$ is Planck’s constant divided by 2π , and $\Delta\mu$ is the chemical-potential⁵ difference across the junction. For a fixed $\Delta\mu$, $\Delta\phi$ evolves in time as $\Delta\phi = -\Delta\mu t/\hbar$, and the corresponding current oscillates as $I = I_c \sin(-\Delta\mu t/\hbar)$. The chemical-potential difference (per Cooper pair) between two superconductors biased at voltage difference ΔV is given by $\Delta\mu = 2e\Delta V$ where e is the charge on the electron, and the corresponding Josephson frequency $f_j = 2e\Delta V/h = 4.82 \times 10^{14} \text{ Hz V}^{-1}$.

Equations (1) and (2), often called the d.c. and a.c. Josephson relations respectively, were re-derived in a more general context by Josephson, Anderson and Feynman^{1–3} who realized that the equations were applicable for any two, weakly coupled, phase-coherent systems. The dramatic aspect of the superconducting case is that the two coupled systems are macroscopic rather than microscopic.

For several decades physicists have recognized a close analogy between superconductors and superfluid helium⁶ in that they are both systems where a large number of particles occupy a single coherent quantum state. It has therefore been natural to search for quantum current oscillations in superfluids. The required condi-

tions for the oscillations to exist in superfluids are that two samples of superfluid be separated by a region small enough so that the wavefunctions of each part can weakly penetrate and overlap that of the other. An example of this would be two reservoirs of superfluid helium in contact through a small aperture. Both the diameter and the length of the aperture must be comparable to the superfluid healing length⁷. This is the minimum length scale over which the magnitude of the superfluid wave function is allowed to vary for

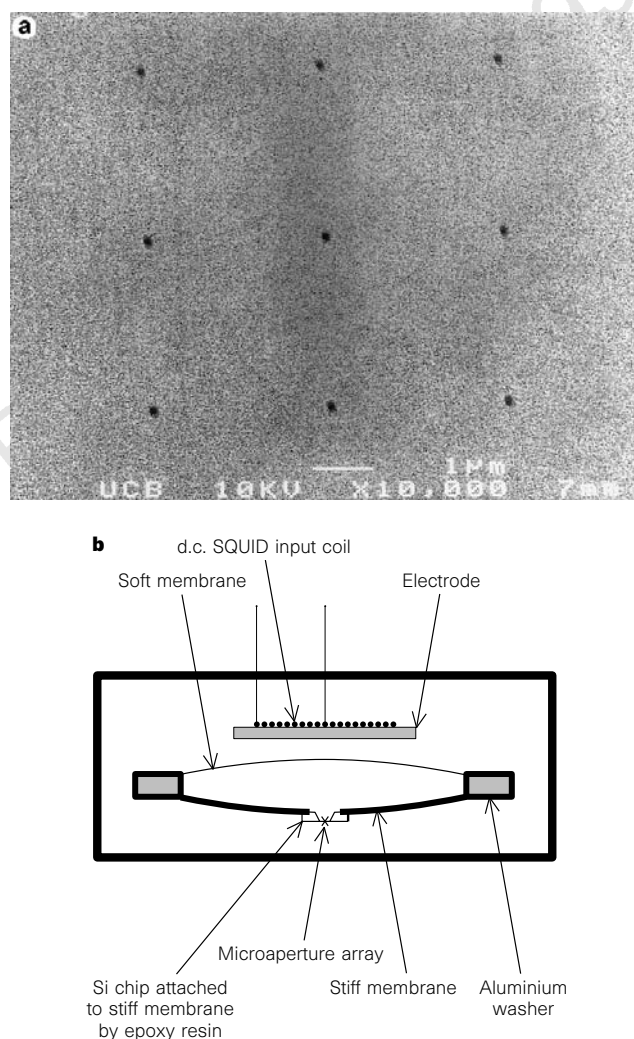


Figure 1 a, Scanning electron micrograph of a small area of the array of apertures in a silicon nitride (SiN) membrane whose thickness is 50 nm. It shows the 3- μm separation between each aperture, and that the aperture diameter is near 100 nm. Larger-area scans showing the whole array confirm that there are 4,225 holes. Scale bar at bottom, 1 μm . **b**, Schematic diagram of the experimental cell. The pill-box-shaped inner cell consists of an 140- μm -thick aluminium washer, to which are attached (by epoxy resin) a stiff diaphragm on the lower side, and a very soft ($1140 \pm 170 \text{ N m}^{-1}$) diaphragm on the top. The Si chip, containing the SiN membrane through which the array of apertures has been etched, is glued in position (by epoxy resin) over a small hole in the lower membrane. The upper membrane is metallized on the outside by evaporating a 100-nm layer of Pb, and then a 20-nm layer of In, onto its surface. It is at the same potential as the Al washer. About 100 μm above the upper membrane is a thin metallized electrode. Changes in the potential of this electrode relative to the membrane are used to apply forces to the soft membrane. Approximately 10 μm behind this electrode are the superconducting wires which form the input coil of the d.c. SQUID. This system is used as the sensitive position sensor. This whole assembly is immersed in a container of superfluid ^3He in contact with a nuclear demagnetization refrigerator.

* Permanent address: Institute for High Pressure Physics, Russian Academy of Sciences, Russia.

reasons of energy minimization. We refer to such an aperture as a weak link.

The equations which are the superfluid equivalent of the superconducting Josephson equations, have the same form as equations (1) and (2). The chemical-potential difference per particle on the right side of equation (2) becomes $\Delta P m / \rho$. Here ΔP is the pressure difference across the weak link, ρ is the liquid density, and m is the mass of either the ^4He atom or twice the ^3He atomic mass, depending on whether one is describing ^4He or ^3He . (We are neglecting a small temperature term in the chemical potential because it is almost negligible in the experiments reported here.) In equation (1) the current, I , now refers to a mass current rather than an electric current. The pressure difference should therefore be accompanied by a mass supercurrent oscillating at the frequency

$$f = \frac{\Delta P m}{\rho h} \quad (3)$$

which is 69 kHz Pa^{-1} for ^4He , and $183.7 \text{ kHz Pa}^{-1}$ for ^3He . We emphasize that this effect does not appear in ordinary fluids, or even for superfluids flowing through tubes which are not weak links. In the weak link, the fluid's response to an applied pressure difference is not simply to accelerate in proportion to that pressure gradient, but rather to oscillate in place at a frequency which is proportional to the pressure difference.

Demonstration of superfluid effects relying on these equations has been very elusive⁸. The fundamental problem is that a superfluid weak link requires an aperture whose dimensions are on the scale of the superfluid healing length, which is only $\sim 0.1 \text{ nm}$ for ^4He and $\sim 50 \text{ nm}$ for ^3He . To observe quantum oscillation phenomena one

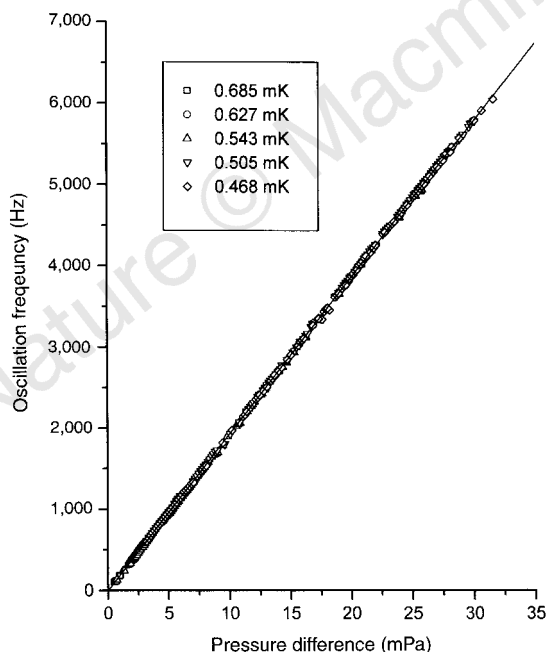


Figure 2 A graph of the frequency of the quantum oscillation detected at the top membrane as a function of ΔP , the pressure difference across the weak-link array. Data for five different temperature T (0.468, 0.505, 0.543, 0.627, 0.685 mK) are plotted. A line is fitted to all the data points, as shown. The slope of the line is 194 kHz Pa^{-1} with a statistical error of $\pm 1 \text{ kHz Pa}^{-1}$, and a systematic uncertainty of $\pm 15 \text{ kHz Pa}^{-1}$. The larger systematic error comes from uncertainties in the capacitance between the soft membrane and the electrode above it. The excellent linear relationship between f and ΔP , and the measured value of the slope (which is predicted to be $183.7 \text{ kHz Pa}^{-1}$), provide a very clear confirmation of equation (3) over more than two orders of magnitude of pressure. For the data at each individual temperature, the f versus ΔP slope has been measured, and the differences between these measured slopes are $<0.5\%$

must first solve two technical problems; the fabrication of such small structures, and the measurement of the extremely small mass currents that would flow through such an aperture. Although these problems are not yet completely solved for ^4He (ref. 9), recent advances have succeeded in both goals for ^3He . Micro-fabrication techniques have led to the development of apertures which are submicrometre in all three dimensions¹⁰. At the same time, displacement transducer development¹¹ (for gravity-wave detection) has led to technology that enables one to monitor the motion of a fluid-driving piston¹², with sensitivity sufficient to detect the small currents predicted^{13–18} in a superfluid weak link. These technical advances have set the stage for observation of Josephson effects in ^3He , and for study of the rich physical phenomena analogous to those observed in superconductors.

There is already one positive experimental indication^{19,20} of a ^3He superfluid current–phase relation similar to equation (1). In that work, the superfluid ^3He mass flow was through a single submicrometre aperture connected in parallel with a larger tube. The authors of references 19 and 20 report that, over a limited temperature range, they can numerically simulate the observed response of their nonlinear superfluid ^3He oscillator, by using equations of motion that incorporate equations (1) and (2) into the dynamics of the system.

Encouraged by these reports, we have pursued the development of a ^3He weak link^{21,22}. Our device uses a square array (Fig. 1a) of 4,225 apertures spaced by 3,000 nm. Each aperture is 100 nm in diameter, in the same 50-nm-thick, silicon nitride membrane¹⁰. We use this ^3He weak link in a temperature range where the superfluid healing length is greater than, or at least comparable to, the aperture diameter. We designed the array with the hope that the apertures would act coherently and have no interference between adjacent apertures, so that the array should behave like a single aperture but with 4,225 times greater current. One of the significant discoveries implied by the results reported below is the fact that the array does indeed behave in this desired manner.

The silicon chip containing the aperture array is placed in the wall of a pill-box-shaped cell, whose top and bottom surfaces are plastic membranes (Fig. 1b). The array is located in a lower, stiff membrane and the position of the upper, soft membrane is monitored with a sensitive displacement detector which can resolve motion as small as $10^{-14} \text{ m Hz}^{-1/2}$. The upper membrane can be manipulated by the application of bias voltages between its conducting surface and a fixed electrode. The cell has an internal volume of $\sim 10 \text{ mm}^3$, and takes several hours to fill by flow of normal liquid ^3He through the array at temperature near 1 K. This cell is placed in a ^3He -filled enclosure, which is in contact with a nuclear demagnetization refrigerator capable of cooling the ^3He to temperatures well below its superfluid transition temperature ($T_c = 0.929 \text{ mK}$).

We apply a step increase in bias potential to the soft membrane, which pulls it abruptly toward the fixed electrode. This creates a stepwise change in the pressure difference ΔP across the array, from zero initial pressure to a value determined by the applied voltage. We record the subsequent motion of the membrane's position, $x(t)$, as mass flow through the array allows it to relax towards its new equilibrium position x_f , where ΔP is again zero. The value of $\Delta x \equiv x_f - x(t)$ is a measure of the instantaneous pressure across the aperture. If the aperture array behaves like a quantum weak link, the relaxing pressure difference $\Delta P(t)$ should be accompanied by mass-current oscillations within the aperture whose frequency is determined by equation (3).

A typical flow transient shows a relaxation of $x(t)$ towards equilibrium. Owing to vibration noise in the displacement transducer, an oscilloscope trace (of the instantaneous voltage which is proportional to x) exhibits no remarkable structure suggestive of the predicted quantum oscillations. But if the electrical output of the displacement transducer is amplified and connected to audio headphones, the listener makes a most remarkable observation. As the

pressure across the array relaxes to zero there is a clearly distinguishable tone smoothly drifting from high to low frequency during the transient, which lasts for several seconds. This simple observation marks the discovery of coherent quantum oscillations between weakly coupled superfluids.

To quantify this assertion we have processed several digitally recorded transient signals. Each transient is broken into equal, short time intervals. The average membrane position during the interval is used to calculate the pressure difference. For each time interval we compute the power spectrum of the membrane motion. As the frequency is drifting during any interval, the peak in the Fourier transform is not sharp. We identify the centre frequency of the strongest spectral component in the range of interest. This is the oscillation frequency corresponding to the pressure difference across the weak link during that interval.

Figure 2 plots the oscillation frequency, f , as a function of ΔP for a series of different temperatures. The frequency is found to be proportional to ΔP , the proportionality factor being 194 kHz Pa^{-1} , with a statistical error of $\pm 1 \text{ kHz Pa}^{-1}$, and a systematic uncertainty of $\pm 15 \text{ kHz Pa}^{-1}$. The linear relationship between f and ΔP , and the measured value of the slope, are in clear agreement with equation (3), over a pressure range of more than two orders of magnitude. This is a direct confirmation of the quantum-oscillation equation for superfluid ^3He .

The relation between frequency and pressure difference is observed to be independent of temperature, to within the error of the measurement. It is expected^{13,14,17} that the actual current-phase relation might become increasingly distorted from the simple sine form in equation (1) as the coherence length gets smaller with falling temperature⁷. The distorted $I(\Delta\phi)$ would lead to higher harmonics of the quantum oscillations. We have searched for these higher harmonics in the spectral signal of the sound, and find that (to within our signal-to-noise ratio) they are absent over a wide range of temperatures.

The experimental results reported here lead to three significant conclusions. (1) The aperture array behaves like a single coherent superfluid weak link. (2) The current-phase relation is similar to the d.c. Josephson equation (equation (1)), that is periodic with modulus 2π . (3) The time evolution of the quantum phase across the array leads to current oscillations that drive the upper diaphragm at the frequency predicted by equation (3).

The remarkable properties of the weak-link array may lead to the observation of several physical phenomena analogous to those found in superconductivity. These include Shapiro²³ steps and, perhaps most significantly, the development of the superfluid analogue of a d.c. SQUID²⁴. □

Received 5 June; accepted 30 June 1997.

- Josephson, B. D. Possible new effects in superconductive tunneling. Thesis, Cambridge Univ. (1962).
- Anderson, P. W. in *Lectures on the Many-Body Problem* (ed. Caianello, E. R.) 113–135 (Academic, New York, 1964).
- Feynman, R. P. *The Feynman Lectures on Physics* Sect. 21-9 (Addison Wesley, New York, 1965).
- Josephson, B. D. Possible new effects in superconductive tunnelling. *Phys. Lett.* **1**, 251–253 (1962).
- Reif, F. *Fundamentals of Statistical and Thermal Physics* Sect. 8-7 (McGraw-Hill, New York, 1965).
- Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* (Hilger, New York, 1990).
- Vollhardt, D. & Woolfe, P. *The Superfluid Phases of Helium-3* Sect. 7-2 (Taylor & Francis, New York, 1990).
- Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* Sect. 7-6 (Hilger, New York, 1990).
- Pereverzev, S. V. Toward the Josephson effect in superfluid helium: how to make 20 nm holes in a 10 nm thick membrane. *J. Low Temp. Phys.* **101**, 573–579 (1995).
- Amar, A., Sasaki, Y., Lozes, R. L., Davis, J. C. & Packard, R. E. Fabrication of submicron apertures in thin membranes of silicon nitride. *J. Vac. Sci. Technol. B* **11**, 259–262 (1993).
- Paik, H. J. Superconducting tunable-diaphragm transducer for sensitive acceleration measurements. *J. Appl. Phys.* **47**, 1168–1178 (1976).
- Avenel, O. & Varoquaux, E. in *Proc. 10th Int. Cryogenic Eng. Conf.* (eds Klipping, G. & Klipping, L.) 587–589 (Butterworths, Guildford, 1986).
- Monien, H. & Tewordt, L. Theory of Josephson flow oscillations in superfluid ^3He -B. *J. Low Temp. Phys.* **62**, 277–300 (1986).
- Hook, J. R. Current-phase relation for flow of an s-wave paired superfluid through a cylindrical channel of finite length. *Jpn. J. Appl. Phys.* **26**, 159–160 (1987).
- Kurkijärvi, J. Josephson current through a short and narrow orifice in a p-wave superfluid. *Phys. Rev. B* **38**, 11184–11187 (1988).
- Thuneberg, E. V. Two dimensional simulation of the Josephson Effect in superfluid ^3He . *Europhys. Lett.* **7**, 441–446 (1988).
- Ullah, S. & Fetter, A. L. Superfluid ^3He -B in a weak link. *Phys. Rev. B* **39**, 4186–4190 (1989).

- Thuneberg, E. V., Kurkijärvi, J. & Sauls, J. A. Quasiclassical theory of the Josephson effect in superfluid ^3He . *Physica B* **165&166**, 755–756 (1990).
- Avenel, O. & Varoquaux, E. Josephson effect and quantum phase slippage in superfluids. *Phys. Rev. Lett.* **60**, 416–419 (1988).
- Varoquaux, E., Avenel, O., Ihas, G. & Salmelin, R. Phase slippage in superfluid ^3He -B. *Physica B* **178**, 309–316 (1990).
- Steinhauer, J. The current-pressure relation for a superfluid ^3He weak link. Thesis, Univ. California, Berkeley (1995).
- Mukharsky, Yu. M., Loshak, A., Schwab, K., Davis, J. C. & Packard, R. E. Study of an array of superfluid ^3He weak links. *Czech. J. Phys.* **46**, 115–116 (1996).
- Shapiro, S. Josephson currents in superconducting tunnelling; the effect of microwaves and other observations. *Phys. Rev. Lett.* **11**, 80–82 (1963).
- Packard, R. E. & Vitale, S. Principles of superfluid-helium gyroscopes. *Phys. Rev. B* **46**, 3540–3549 (1992).

Acknowledgements. In early stages of this work we were helped by J. Steinhauer, K. Schwab, Yu. M. Murkarsky and I. Schuster. E. W. Hudson made important contributions to the data acquisition system. Figure 1b was made by R. Orr. The aperture array was constructed at the Berkeley Microfabrication Laboratory. This work was supported in part by the National Science Foundation, the Office of Naval Research, and the Packard Foundation.

Correspondence and requests for materials should be addressed to R.E.P. (e-mail: packard@socrates.berkeley.edu).

Graphitic cones and the nucleation of curved carbon surfaces

A. Krishnan*, E. Dujardin†, M. M. J. Treacy*, J. Hugdahl‡, S. Lynum‡ & T. W. Ebbesen*

* NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA

† Laboratoire de Chimie, Collège de France, 11, place M. Berthelot, 75005 Paris, France

‡ Kvaerner Engineering a.s., 1324 Lysaker, Norway

§ ISIS, Louis Pasteur University, rue Blaise Pascal, 67000 Strasbourg, France

The nucleation and growth of curved carbon structures, such as fullerenes, nanotubes and soot, are still not well understood. A variety of models have been proposed^{1–17}, and it seems clear that the occurrence of pentagons, which yield 60° disclination defects in the hexagonal graphitic network, is a key element in the puzzle. The problem of nucleation has been complicated by the great variety of structures observed in any one sample. Here we report an unusual carbon sample generated by pyrolysis of hydrocarbons, consisting entirely of graphitic microstructures with total disclinations that are multiples of $+60^\circ$. The disclination of each structure corresponds to the presence of a given number of pentagons in the seed from which it grew: disks (no pentagons), five types of cones (one to five pentagons), of which only one was known previously¹⁸, and open tubes (six pentagons). Statistical analysis of these domains shows some unexpected features, which suggest that entropy plays a dominant role in the formation of disclinations. Furthermore, the total disclination of a domain is determined mainly at the nucleation stage.

Several models have been proposed for the formation of fullerenes, which are also relevant to the nucleation of other curved structures such as nanotubes. The “pentagon-road”^{1,3} and the “ring-stacking” models⁴ are among the best known. The essence of the pentagon-road model is that small flat hexagonal networks of carbon have so much dangling-bond energy at the edges that, if the temperature and timescale are such as to permit annealing, the structures will try to eliminate dangling bonds by folding and incorporating pentagons until the structure is closed (12 pentagons or 4π total disclination). Dynamic calculations support this idea⁵. Reduction of the reaction enthalpy ΔH is the overriding concern in the pentagon-road model. In the ring-stacking model, it is proposed that curved surfaces are created by the selective assembly of monocyclic carbon rings⁴, to account for the fact that only very few of the possible fullerene isomers are formed¹⁵. As it considers only certain reaction paths, the ring-stacking model is essentially driven by entropy changes (ΔS). Monocyclic rings are known to be the

dominant structures for cluster sizes between 10 and 30 carbon atoms, and some elegant experiments show that they can lead to the formation of fullerenes^{9–11}, although not necessarily in such a way as to support the ring-stacking model.

Although they start from very different considerations, both models contain as their key the probability of forming a disclination. However, we know next to nothing about the probability of forming curved carbon surfaces having disclinations consistent with one, two, three pentagons and so on. Furthermore, what does this probability tell us about the products and their yields?

Our sample was made by the pyrolysis of hydrocarbons in a carbon arc. We use an industrial-scale carbon-arc plasma generator (torch configuration) under a continuous flow of hydrocarbon (in this case heavy oil) at a typical feed rate of 50–150 kg h⁻¹ and a reactor pressure of 2–3 bar. The solid carbon products are recovered and the residual undecomposed hydrocarbon and hydrogen are fed back into the reactor. The effective plasma temperature is hard to determine but from the level of graphitization of the sample, we estimate it to be at least 2,000 °C. The sample is composed nearly entirely of turbostratic (where the graphitic sheets are not ordered relative to each other in the plane) microstructures (Fig. 1) which have total disclinations (TD) that are a multiple of 60°: $TD = P \times 60^\circ$, where $P \geq 0$, and corresponds to the effective number of pentagons necessary to produce that particular total disclination. Besides disks ($P = 0$), there are also a variety of cones. By considering the symmetry of a graphite sheet and Euler's theorem, it can be shown that only five types of cone can be made from a continuous sheet of graphite corresponding to values of P between 1 and 5. These are all present in this sample (Fig. 2a–e). The cone angle θ is given by $\sin(\theta/2) = 1 - (P/6)$. As far as we know, only the cone with $TD = 300^\circ$ ($P = 5$) has been observed before¹⁸. Nanotubes are also present, although in smaller amounts. Figure 2f shows how the parallel basal planes follow the contour of the cone tip. Cones and nanotubes form ~20% of the sample, the rest being mainly disks.

We measured ~1,700 microstructures from transmission electron micrographs. The cone angle θ can be deduced from the projected dimensions of the tilted cones. A histogram of frequency versus cone angle θ (Fig. 3a) reveals that there are a number of maxima, which become better defined at smaller angles. Each maximum corresponds to a total disclination equivalent to between no pentagons and six pentagons. The width of the peaks increases with the cone angle θ mainly owing to measurement errors, it being hardest to accurately de-project the largest cone angles. In Fig. 3b the same data are reduced to the total count found at the average angle of each peak, with error bars, and corrected for underestimation of occurrence of cones with large angles. These angles correspond closely to those expected for solid angles which are multiples of 60° or an integer number of pentagons (open circles).

Figure 3 gives us the probability of nucleating each of type of microstructure and, therefore, of having seed structures with a given disclination (or number of pentagons). Surprisingly, there is a broad peak with a maximum at a cone angle corresponding to three pentagons. Reasoning on purely enthalpic grounds, one would have expected that small disclinations would have been more probable as they are less strained¹⁹. Thus, a Boltzmann distribution cannot account for the observed distribution as shown in Fig. 3b.

This result indicates that, whereas enthalpy (ΔH) is obviously the driving force for the formation of curved structures, entropy (ΔS) plays a dominating role in the distribution of disclinations formed. Here one must understand entropy changes in terms of the multitude of chemical pathways leading to a particular disclination. Furthermore, temperature will moderate the various possible pathways. A model like the ring-stacking model considers only a small subset of all the possible pathways leading to a given disclination. Monocyclic ring structures have been proved to favour the formation of fullerenes^{9–11}. However, it is not clear if these monocyclic rings

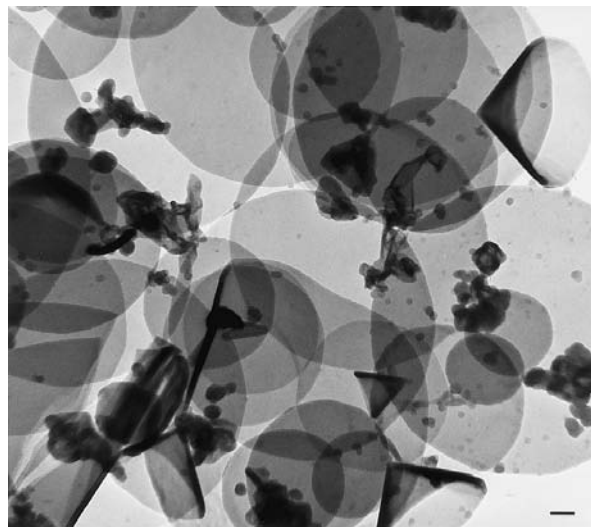


Figure 1 Transmission electron micrograph of the microstructures in the sample. Scale bar, 200 nm.

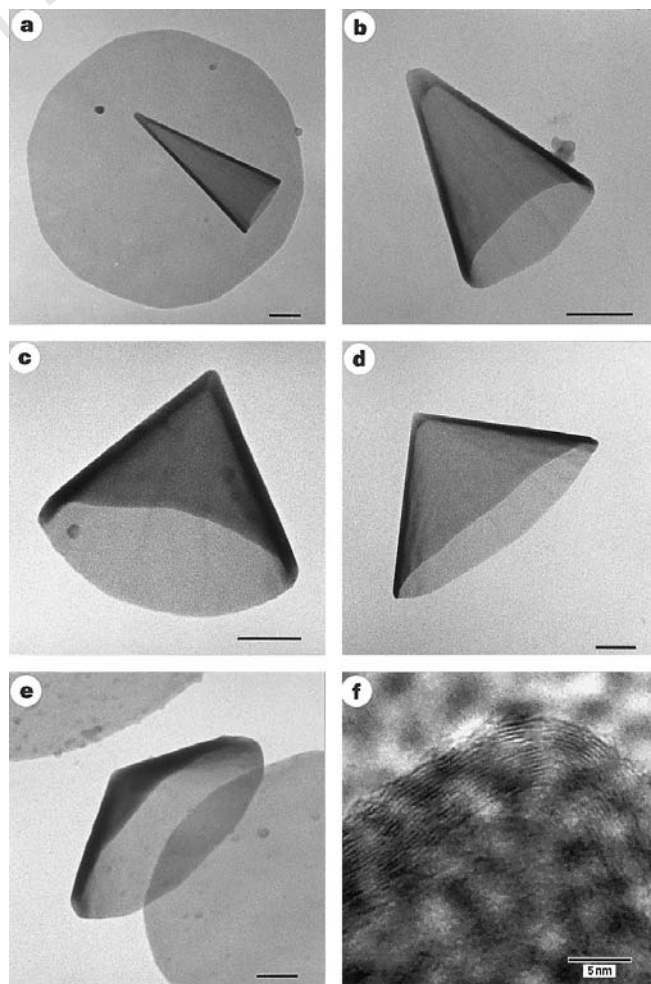


Figure 2 Examples of the five types of cones in the sample (scale bars in a–e, 200 nm): **a**, $TD = 300^\circ$, $\theta = 19.2^\circ$; **b**, $TD = 240^\circ$, $\theta = 38.9^\circ$; **c**, $TD = 180^\circ$, $\theta = 60.0^\circ$; **d**, $TD = 120^\circ$, $\theta = 84.6^\circ$; and **e**, $TD = 60^\circ$, $\theta = 112.9^\circ$. Note the faceting of the disk in **a, f**. Higher magnification image of a cone tip (scale bar, 5 nm).

form the dominant pathway for the generation of all the disclinations observed in curved carbon structures. In principle, cones can be built by the growth of graphitic structures onto such rings. From statistical-mechanics considerations, there are many more ways of embedding a ring of a given size around the tip of a conical graphitic surface than into a graphite sheet. For example, there are 16 ways to embed a C_{16} ring into a planar graphite sheet about a fixed point, whereas there are 2,500 conformations of the same ring that will wrap around the tip of a $P = 2$ cone. Furthermore, starting from an odd-numbered open ring, a structure where P is even cannot be

grown and vice versa, unless crosslinking of the ring occurs. Consequently, the pentagons are not the sole source of disclinations in the system. A disclination is inherent to the monocyclic ring topology and becomes fixed once growth has started. With monocyclic carbon seeds, pentagons are a consequence of the disclination, not vice versa. Of course, other seed topologies such as crosslinked rings will also be important. We are exploring this entropic "monocyclic avenue" further.

Although a mixture of hydrocarbons was the source of carbon in these experiments, hydrocarbons are known to dehydrogenate at high temperatures², yielding under the right conditions the same family of products as a pure carbon plasma. Therefore, these findings should have broad implications for high-temperature generation of curved carbon structures. The wide distribution of disclinations, and the fact that this distribution is controlled by entropy, indicate that, in the absence of catalysts, it will always be hard to make a single type of disclination (except totally spheroidal structures) from a soup of carbon clusters, as discussed next.

Nanotubes and nanoparticles formed in a carbon arc have a bimodal distribution in terms of length^{13,20}. Furthermore, yields of nanotubes versus nanoparticles have never been higher than about half the sample, despite numerous efforts over several years to increase the yields. Those results can now be better understood in the light of the distribution of seeds. At the higher temperature (3,700 °C) of formation of nanotubes in the arc, the probability of forming pentagons in a seed increases⁶. The optimal yield of nanotubes versus nanoparticles indicates that the distribution of seeds must be centred near six pentagons. However, all seeds with seven or more pentagons will inevitably form closed nanoparticles as the edges of the curved surfaces are on collision course as they grow. Seeds with six or fewer pentagons have a finite probability of becoming tubes, depending on the kinetics of growth and the probability of the introduction of another pentagon in the structure during the growth stage (see below). So what was originally a wide distribution of disclinations is now narrowed into mainly two products, hence the bimodal distribution²⁰. The broad distribution of seeds also implies that it would be difficult to improve the yield of nanotubes beyond what has already been achieved. Conversely it is easy to reduce the yield of nanotubes by shifting the disclination maximum of the seeds, as found in nanotube fabrication.

The formation of spheroidal soot particles from hydrocarbons involves both nucleation and growth stages¹⁶. Although very little is known of the nucleation process¹⁷, the possible connection to fullerene formation has been pointed out^{1,2}. By analogy with colloidal suspensions, liquid hydrocarbon nuclei have been invoked to explain the formation of uniformly spherical soot particles¹⁶. In the context of these results, just like the nanoparticles, it suffices that the distribution of disclinations in the seeds is such that most of them have disclinations $>360^\circ$ for all them to result in uniform spherical or polygonal particles. In other words, a broad distribution of disclinations in the seeds leads to nearly uniformly spheroidal soot particles.

Our results indicate that the total disclination, or curvature, is nearly always determined in the nucleation stage. A change in disclination after seeding (for example, the addition of pentagons or heptagons during growth) was rarely found in these samples. An example of such occurrence is shown in Fig. 4 where a cone with a 300° disclination has turned into a tube by the addition of a sixth pentagon. The relatively low probability of the formation of pentagons or heptagons during growth could point to different mechanisms of disclination formation in the nucleation and growth stages. Alternatively, it could be due simply to chemical factors during growth such as differences in rates or the thickening of structures. □

Received 21 April; accepted 17 June 1997.

1. Zhang, Q. L. *et al.* Reactivity of large carbon clusters: spheroidal carbon shells and their possible relevance to the formation and morphology of soot. *J. Phys. Chem.* **90**, 525–528 (1986).
2. Kroto, H. W. & McKay, K. The formation of quasi-icosahedral spiral shell carbon particles. *Nature* **331**, 328–331 (1988).

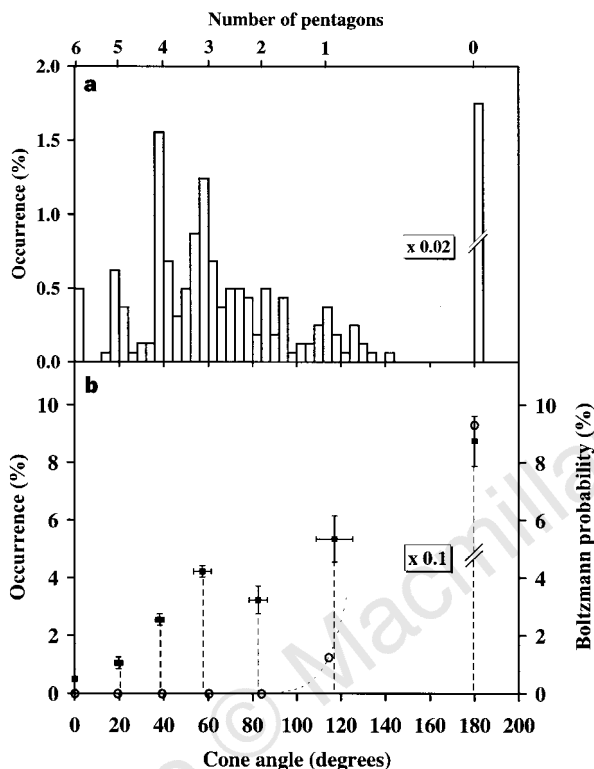


Figure 3 **a**, Statistical distribution of the measured cone angles among the microstructures; **b**, averaged distribution centred at the seven possible disclinations. The open circles and dotted line represent the calculated Boltzmann distribution of microstructures for a temperature of 2,000 °C and a pentagon energy of formation of 0.75 eV (ref. 19).

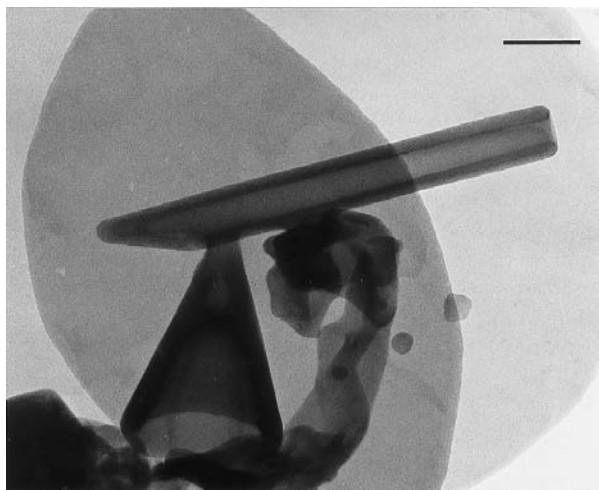


Figure 4 Example of tube formation after nucleation of a cone with a 300° disclination (scale bar, 200 nm). Note the open-ended growth.

3. Smalley, R. E. Self-assembly of the fullerenes. *Acc. Chem. Res.* **25**, 98–105 (1992).
4. Wakabayashi, T. & Achiba, Y. A model for the C_{60} and C_{70} growth mechanism. *Chem. Phys. Lett.* **190**, 465–468 (1992).
5. Robertson, D. H., Brenner, D. W. & White, C. T. On the way to fullerenes: molecular dynamics study of the curling and closure of graphitic ribbons. *J. Phys. Chem.* **96**, 6133–6135 (1992).
6. Endo, M. & Kroto, H. W. Formation of carbon nanofibers. *J. Phys. Chem.* **96**, 6941–6944 (1992).
7. Iijima, S., Ajayan, P. M. & Ichihashi, T. Growth model for carbon nanotubes. *Phys. Rev. Lett.* **69**, 3100–3103 (1992).
8. Saito, Y., Yoshikawa, T., Inagaki, M., Tomita, M. & Hayashi, T. Growth and structure of graphitic tubules and polyhedral particles in arc-discharge. *Chem. Phys. Lett.* **204**, 277–282 (1993).
9. McElvany, S. W., Ross, M. M., Goroff, N. S. & Diederich, F. Cyclocarbon coalescence: mechanisms for tailor-made fullerene formation. *Science* **259**, 1594–1596 (1993).
10. von Helden, G., Gotts, N. G. & Bowers, M. T. Experimental evidence for the formation of fullerenes by collisional heating of carbon rings in the gas phase. *Nature* **363**, 60–63 (1993).
11. Hunter, J., Fye, J. & Jarrold, M. F. Carbon rings. *J. Phys. Chem.* **97**, 3460–3462 (1993).
12. Harris, P. J. F., Tsang, S. C., Claridge, J. B. & Green, M. L. H. High resolution microscopy studies of a microporous carbon produced by arc-evaporation. *J. Chem. Soc. Faraday Trans.* **90**, 2799–2802 (1994).
13. Gamaly, E. G. & Ebbesen, T. W. Mechanism of carbon nanotube formation in the arc discharge. *Phys. Rev. B* **52**, 2083–2089 (1995).
14. Charlier, J.-C., De Vita, A., Blase, X. & Car, R. Microscopic growth mechanisms for carbon nanotubes. *Science* **275**, 646–649 (1997).
15. Achiba, Y. et al. in *The Chemical Physics of Fullerenes 10 (and 5) Years Later* (ed. Andreoni, W. 139–47 (Kluwer Academic, Dordrecht, 1996).
16. Lahaye, J. & Prado, G. in *Particulate Carbon* (eds Siegl, D. C. & Smith, G.W.) 33–55 (Plenum, New York, 1981).
17. Harris, S. J. & Weiner, A. M. Chemical kinetics of soot particle growth. *Annu. Rev. Phys. Chem.* **36**, 31–52 (1985).
18. Ge, M. & Sattler, K. Observation of fullerene cones. *Chem. Phys. Lett.* **220**, 192–196 (1994).
19. Ihara, S., Itoh, S., Akagi, K., Tamura, R. & Tsukada, M. Structure of polygonal defects in graphitic carbon sheets. *Phys. Rev. B* **54**, 14713–14719 (1996).
20. Ebbesen, T. W. in *Carbon Nanotubes: Preparation and Properties* (ed. Ebbesen, T.W.) Ch. 4 (CRC, Boca Raton, 1997).

Acknowledgements. We thank M. E. Bisher for technical assistance.

Correspondence and requests for materials should be addressed to T.W.E.

Photoisomerization in dendrimers by harvesting of low-energy photons

Dong-Lin Jiang & Takuzo Aida

Fields and Reactions, Precursory Research for Embryonic Science and Technology, Japan Science and Technology Corporation, and Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Infrared radiation can induce low-frequency molecular vibrations, but, with the exception of hydrogen-bond reorganization^{1–3}, the excitation energy tends to be dissipated rapidly through molecular collisions rather than inducing photochemical changes. Here we show that in a macromolecular system that is designed to be insulated against collisional energy scattering, infrared absorption can excite photoisomerization by multiphoton intramolecular energy transfer. We have prepared highly branched dendrimers^{4–6} from aryl ethers with a photoisomerizable azobenzene core, in which infrared excitation of the aromatic units is apparently followed by a channelling of the absorbed energy to the core while the dendrimer matrix protects against collisional de-excitation. These findings suggest a strategy for harvesting low-energy photons to effect chemical transformations.

Dendrimers are nanometre-size macromolecules with a regular tree-like array of branch units^{4–6}. We have recently synthesized dendrimer porphyrins in which a porphyrin chromophore is spatially isolated by the dendrimer framework^{7–10}. As may be expected from the gradient in branch density from the interior to the exterior, the surface of these large dendrimers is conformationally frozen, whereas the interior space is not constrained irrespective of the size of the dendrimer framework, so that the core porphyrin is able to undergo free conformational motion⁸. Thus spherical, large dendrimers should provide a non-constrained environment insulated from collisional energy scattering. This motivated us to explore the possibility of intramolecular energy

transfer within these unimolecular matrices.

For this purpose, we have synthesized a series of azobenzene-containing aryl ether dendrimers (*trans*-LnAZO, where *n* is the number of aromatic layers and is 1, 3, 4 or 5; Fig. 1), by alkaline-mediated coupling of dendritic benzyl bromides (L(*n* – 1)Br; ref. 8) with *trans*-3,3',5,5'-tetrahydroxyazobenzene. We characterized this series of *trans*-LnAZO compounds unambiguously, by using ¹H NMR, ultraviolet–visible spectroscopy and matrix-assisted laser desorption ionization time-of-flight mass spectroscopy (further details are available; see Supplementary Information). Azobenzenes are photochromic molecules that isomerize from *trans* to *cis* on ultraviolet irradiation, and from *cis* to *trans* either thermally or on exposure to visible light¹¹. Single-photon (ultraviolet) photoisomerization of azobenzene dendrimers similar to ours was demonstrated recently¹².

We measured ¹H NMR pulse relaxation times (*T*₁) of the *trans*-LnAZO family at 21 °C in CDCl₃, and confirmed an egg-like structural resemblance⁸ to higher-generation *trans*-LnAZOs. The *T*₁ value for the methoxy protons on the exterior surface (filled circles in Fig. 2a) decreased sharply as the dendrimer framework became larger whereas that of the protons in the interior azobenzene functionality (open circles in Fig. 2a) remained almost intact. Thus L4AZO and L5AZO possess a non-constrained interior environment within a stiff exterior shell.

On ultraviolet radiation at 21 °C in CHCl₃, *trans*-LnAZO (*n* = 1, 3–5, 5×10^{-5} M) isomerized normally to *cis*-LnAZO, which then gradually isomerized back to the *trans* form after the irradiation was stopped. When a CHCl₃ solution of *cis*-L5AZO in a KBr cell was exposed to infrared radiation (a 75-W glowing nichrome source) in an infrared spectrophotometer, it isomerized very rapidly to *trans*-L5AZO. We found that this infrared-radiation-induced isomerization depends greatly on the size of the dendrimer framework (Fig. 2b): on exposure to the infrared radiation *cis*-L5AZO completely isomerized to the *trans* form in only 8 min (Fig. 3b) at a rate constant of $3.4 \times 10^{-3} \text{ s}^{-1}$, which is 260 times as high as that of the thermal isomerization at 21 °C in the dark ($1.3 \times 10^{-5} \text{ s}^{-1}$, Fig. 2b) and even 23 times as high as the rate constant found on irradiation with visible light ($440 \pm 9 \text{ nm}$, 21 °C). Similarly, under infrared illumination *cis*-L4AZO underwent rapid isomerization to *trans*-L4AZO (Fig. 2b). In sharp contrast, we observed no effect of infrared radiation for the isomerization of the smaller homologues *cis*-L1AZO (Fig. 3a) and *cis*-L3AZO. *cis*-L1AZO was also unaffected by the infrared radiation even in the presence of a large dendritic benzyl alcohol (L5OH; 4 equiv.) unanchored to the azobenzene unit. Furthermore, although the molecular weight is almost comparable to that of L4AZO, the isomerization of *cis*-mono-L6'AZO (Fig. 1), a non-spherical mono-dendritic azobenzene, was not facilitated by infrared. Thus spatial isolation of the azobenzene functionality by the highly constrained, large dendrimer matrix (Fig. 2a) seems to be a prerequisite for the infrared-radiation-induced isomerization.

We further investigated the isomerization of *cis*-L5AZO (5×10^{-5} M, in CHCl₃ at 21 °C) by using monochromatized infrared radiation (a 75-W glowing nichrome source through a JASCO model CT-25C monochromator, bandwidth $\pm 50 \text{ cm}^{-1}$) of three different wavenumbers; $2,500 \text{ cm}^{-1}$ (transparent for LnAZO), $1,597 \text{ cm}^{-1}$ (a stretching vibrational band for aromatic rings) and $1,155 \text{ cm}^{-1}$ (a stretching vibrational band for CH₂–O), and found that only the $1,597 \text{ cm}^{-1}$ radiation facilitates the isomerization reaction. This observation indicates that the infrared energy absorbed by the aromatic rings excites the *cis*-to-*trans* isomerization. We consider the following two possibilities for the mechanism of this effect; (1) direct infrared excitation of the interior azobenzene moiety, and (2) intramolecular energy transfer from the dendrimer matrix to the azobenzene. To investigate (2), we selectively excited the dendrimer framework of *cis*-LnAZO at its characteristic 280-nm absorption, where we again observed a definite

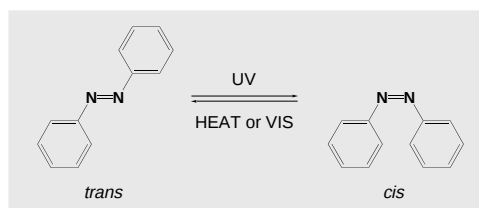


Figure 1 Schematic representations of the isomerization of azobenzene (top left) and the structures of *trans*-L_nAZO (*n* = 1,3-5) and *trans*-mono-L6'AZO; here *n* represents the number of aromatic layers. In the equation at top left, *UV* indicates ultraviolet radiation and *VIS* indicates visible light radiation.

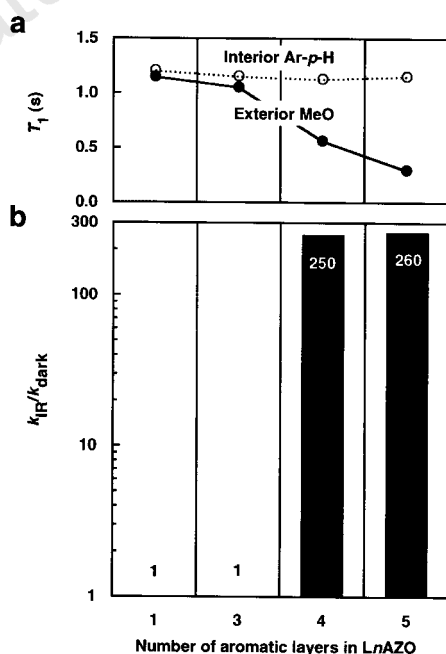
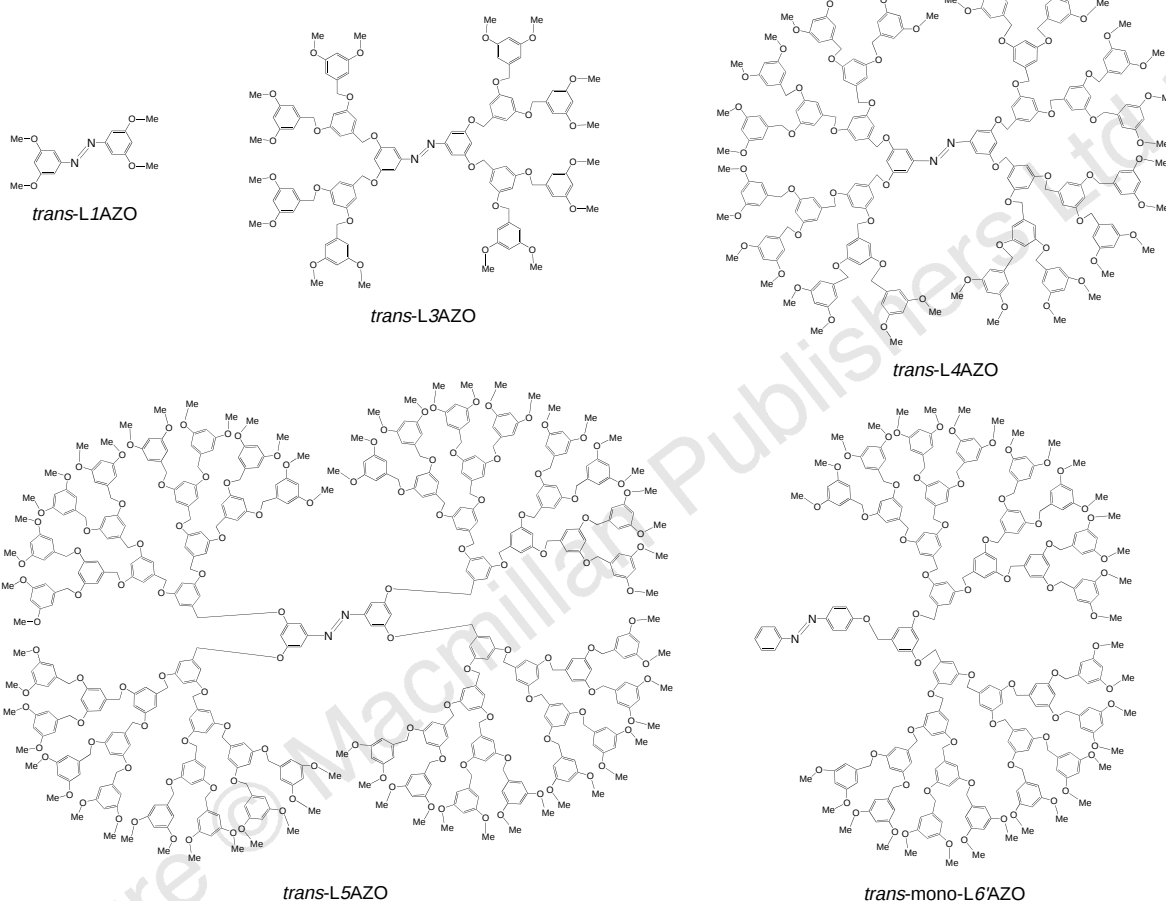


Figure 2 Effect of the number of aromatic layers (*n*) in L_nAZO on: **a**, ¹H NMR pulse relaxation times (*T*₁) of *trans*-L_nAZO (*n* = 1,3-5) at 21 °C; **b**, first-order rate constants of the *cis*-to-*trans* isomerization of L_nAZO (*n* = 1,3-5) at 21 °C under infrared irradiation (*k*_{IR}) relative to those of the thermal isomerization in the dark (*k*_{dark}). The experimental procedures for **b** are described in the Methods section.

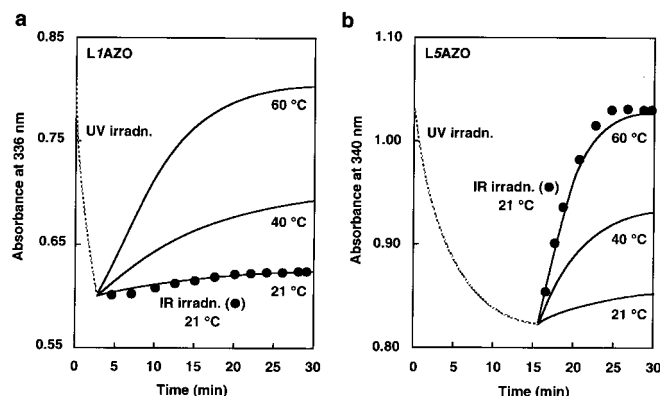


Figure 3 Isomerization profiles of L1AZO (a) and L5AZO (b) in infrared radiation (●) at 21 °C in comparison with the thermal isomerization profiles (solid lines) at 21, 40 and 60 °C in the dark. Experimental details are given in the Methods section.

acceleration of the *cis*-to-*trans* isomerization for L4AZO and L5AZO, but not for L1AZO, L3AZO and mono-L6'AZO (further details are available; see Supplementary Information). In every case, the fluorescence from the dendrimer framework (310 nm) was not quenched at all. Considering a possible internal conversion via the singlet excited state to vibrational modes of the ground state; the correlation of the observed results on 280-nm excitation with those in Fig. 2b indicates that the infrared-radiation-induced isomerization of L_n AZO involves a contribution of the matrix-to-core intramolecular energy transfer. We therefore conclude that the large dendrimer matrix here does not simply serve to insulate the interior units from collisional energy scattering, but it also functions as an efficient photon-harvesting antenna.

For comparison, we investigated thermal isomerization of *cis*-L5AZO at three different temperatures (Fig. 3b) and found that the rate of isomerization on infrared excitation is almost comparable to that of the thermal isomerization at 60 °C. From the data shown in Fig. 3b, the activation free energy ($\Delta G^\ddagger$) at 21 °C for the thermal isomerization of *cis*-L5AZO was found to be 19.4 kcal mol⁻¹ (0.84 eV). We further investigated the effect of different photon fluxes on the infrared-radiation-induced isomerization of *cis*-L5AZO by changing the distance d of the sample from a point light source; the photon flux applied to the sample is in inverse proportion to d^2 . The observed dependence of the isomerization rate constant (k_{IR}) on d^2 (Fig. 4) indicates that 4.9 photons (0.98 eV at 1,597 cm⁻¹) are involved in this photochemical process. L5AZO bears 62 aromatic groups, and may possibly absorb several photons simultaneously even in such a low photon flux field.

There is much interest in the design of molecular systems for light harvesting and efficient conversion of photons into chemical energy^{13–16}. Our results suggest a new strategy for using low-energy photons for chemical transformations, and the possibility of using dendrimers as tunable light-harvesting matrices. □

Methods

A CHCl₃ solution of *trans*- L_n AZO (5×10^{-5} M), in a quartz cell of 10-mm path length connected to a 10-ml round-bottomed flask with a three-way stopcock, was irradiated by a 300-W xenon arc light through a band pass filter (340 ± 9 nm) to induce the *trans*-to-*cis* isomerization (broken lines). After the spectral change subsided, the irradiated solution was transferred to a KBr cell of

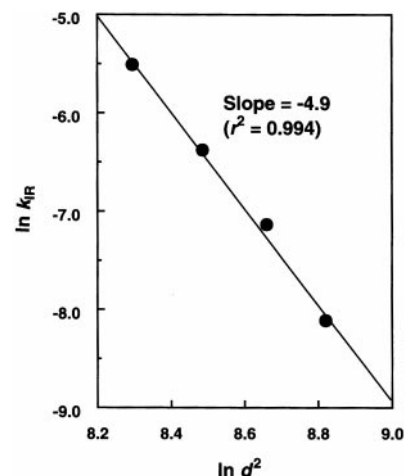


Figure 4 Dependence on the irradiation distance d of the first-order isomerization rate constant (k_{IR}) of *cis*-L5AZO when exposed to infrared radiation. A CHCl₃ solution of *cis*-L5AZO (5×10^{-5} M) in a KBr cell of 5-mm path length was exposed at 21 °C to infrared radiation (a 75-W glowing nichrome point source) at four different distances (63.3, 69.6, 76.0 and 82.3 mm); the rate constants (k) were obtained as given in the Methods section.

5-mm path length and exposed at 21 °C to infrared radiation (a 75-W glowing nichrome source) in an infrared spectrophotometer (JASCO FT-IR model 5300), where the temperature increase was only within 1 °C, as observed by a thermocouple microdetector. The entire solution was periodically transferred to a quartz cell of 10-mm path length and the electronic absorption spectrum was taken, to follow the *cis*-to-*trans* isomerization. For thermal isomerization, the quartz cell, containing the irradiated solution at 340 ± 9 nm, was set in a ultraviolet–visible spectrophotometer equipped with a temperature controller, and the absorbance at 336 nm (for L1AZO) or 340 nm (for L5AZO) was continuously monitored at a designated temperature.

Received 18 November 1996; accepted 17 June 1997.

- Hall, R. T. & Pimentel, G. C. Isomerization of nitrous acid: an infrared photochemical reaction. *J. Chem. Phys.* **38**, 1889–1897 (1963).
- Fredin, L., Nelander, B. & Ribbegard, G. Infrared spectrum of the water dimer in solid nitrogen. *J. Chem. Phys.* **66**, 4073–4077 (1977).
- Scherer, P. O. J., Seilmeier, A. & Kaiser, W. Ultrafast intra- and intermolecular energy transfer in solutions after selective infrared excitation. *J. Chem. Phys.* **83**, 3948–3957 (1985).
- Fréchet, J. M. J. Functional polymers and dendrimers: reactivity, molecular architecture, and interfacial energy. *Science* **263**, 1710–1715 (1994).
- Tomalia, D. A. Starburst/cascade dendrimers: fundamental building blocks for a new nanoscopic chemistry set. *Adv. Mater.* **6**, 529–539 (1994).
- Jansen, J. F. G. A. *et al.* Encapsulation of guest molecules into a dendritic box. *Science* **266**, 1226–1229 (1994).
- Jin, R.-H., Aida, T. & Inoue, S. 'Caged' porphyrin: the first dendritic molecules having a core photochemical functionality. *Chem. Commun.* 1260–1261 (1993).
- Tomoyose, Y. *et al.* Aryl ether dendrimers with an interior metalloporphyrin functionality as a spectroscopic probe: interpenetrating interaction with dendritic imidazoles. *Macromolecules* **29**, 5236–5238 (1996).
- Sadamoto, R., Tomioka, N. & Aida, T. Photoinduced electron transfer reactions through dendrimer architecture. *J. Am. Chem. Soc.* **118**, 3978–3979 (1996).
- Jiang, D.-L. & Aida, T. Dendritic iron porphyrin as a novel haemoprotein mimic: effects of dendrimer cage on dioxygen-binding activity. *Chem. Commun.* 1523–1524 (1996).
- Kumar, G. S. & Neckers, D. C. Photochemistry of azobenzene-containing polymers. *Chem. Rev.* **89**, 1915–1925 (1989).
- Junge, D. M. & McGrath, D. V. Photoresponsive dendrimers. *Chem. Commun.* 857–858 (1997).
- Devadoss, C., Bharathi, P. & Moore, J. S. Energy transfer in dendritic molecules: molecular size effects and the role of an energy gradient. *J. Am. Chem. Soc.* **118**, 9635–9644 (1996).
- Webber, S. E. Photon-harvesting polymers. *Chem. Rev.* **90**, 1469–1482 (1990).
- Fetisova, Z. G. *et al.* Long-range molecular order as an efficient strategy for light harvesting in photosynthesis. *Nature* **334**, 633–634 (1988).
- Yamanaka, G. *et al.* Molecular architecture of a light-harvesting antenna. *J. Biol. Chem.* **255**, 11004–11010 (1980).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank T. Kitamori for discussions, Y. Okamoto and E. Yashima for molecular modelling calculations, and JASCO Co. Ltd. for providing a model CT-25C infrared light monochromator. D.-L.J. thanks JSPS for a Young Scientist Fellowship.

Correspondence and requests for materials should be addressed to T.A. (e-mail: aida@chiral.t.u-tokyo.ac.jp).

Past ultraviolet radiation environments in lakes derived from fossil pigments

Peter R. Leavitt*, Rolf D. Vinebrooke*, David B. Donald†, John P. Smol‡ & David W. Schindler§

* Limnology Laboratory, Department of Biology, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

† Environment Canada, Ecosystems Division, Regina, Saskatchewan, S4P 4K1, Canada

‡ Paleocological Environmental Assessment and Research Laboratory, Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

§ Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Natural levels of ultraviolet (UV) radiation can harm organisms in shallow aquatic ecosystems in which concentrations of photo-protective dissolved organic carbon are low^{1–3}. These compounds can be removed as a result of acidic precipitation and climate changes, an effect which may have recently been manifested in up to 200,000 boreal lakes^{4,5}. Unfortunately, meteorological and biological monitoring studies are usually too brief to record the magnitudes of past changes in UV radiation fluxes and their effects. Here we demonstrate that certain fossil pigments in lake sediments can be used to document historical changes in the UV radiation environment of lakes. These pigments are produced by benthic algae when exposed to UV radiation and show sedimentary concentrations that are correlated to the depth of penetration of UV radiation within lakes. Analysis of fossil profiles from the sediments of two mountain lakes suggests that past UV radiation penetration has sometimes been—at least in these mid-latitude lakes—greater than during the period of anthropogenic stratospheric ozone depletion.

Analysis of sedimentary deposits in lakes has been used to document a wide variety of environmental perturbations, including climate warming, acidic precipitation and lake eutrophication^{6,7}. We have proposed that fossil analyses might also record changes in past UV radiation environments because pigments from algae are often well-preserved in lake sediments⁸, and because some algae produce unique pigments on exposure to UV radiation^{9,10}. Recently we have recorded the presence of UV-radiation-absorbing pigments in the sediments of alpine lakes¹¹ where UV radiation was known to penetrate deeply¹² and inhibit growth of benthic algae on artificial substrates¹³. Mountain lakes may be particularly sensitive to UV radiation both because they are exposed to naturally high fluxes of it^{14,15} and because they contain low concentrations of terrestrial dissolved organic carbon (DOC)¹⁶, the main attenuator of UV radiation in lakes¹⁷.

Surveys of clear lakes were conducted in 1994 and 1995 to establish the distribution of UV-radiation-absorbing sedimentary pigments among lakes of differing depth and DOC content. Lakes (62) in Banff, Jasper and Yoho National Parks in the Canadian Rocky Mountains (51° 00'–53° 30' N, 115° 30'–119° 30' W) were chosen to span a wide range of limnological conditions¹⁸. Ekman grab samples were used to collect sufficient sediments from the deepest point in each lake. Flocs of recently deposited algae were recovered from the mud–water interface, indicating that sediments were relatively undisturbed. Surface sediments (2–3 cm deep) were isolated from samples by hand-held mini-core sampler, frozen and analysed for pigments and organic matter content. Depth-integrated water samples were collected from the epilimnion of each lake in August 1995 and analysed for DOC content. Weekly samples over three summers in several alpine lakes revealed that

mean DOC concentrations ($\pm$ standard error, s.e.) varied little between years ($0.68 \pm 0.05 \text{ mg l}^{-1}$, $n = 3$) or within years (s.e. = 0.10–0.16, $n = 9$ –11) under current climate regimes (D.W.S., unpublished data).

Sedimentary pigments, including carotenoids, chlorophylls and derivatives from algae and bacteria, were used to estimate the

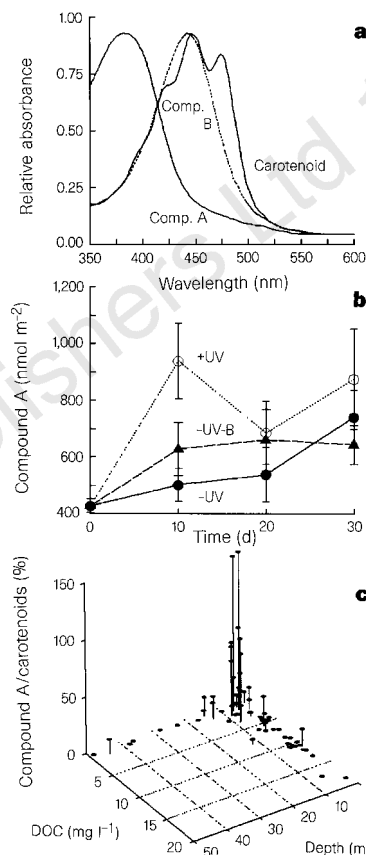


Figure 1 a–c. Absorbance spectra, production and distribution of ultraviolet-radiation-specific pigments isolated from mountain lakes. **a**, Relative absorbance of a carotenoid from green algae (lutein; absorption peaks of 472, 444.5, 420 nm), compound B ($\lambda_{\text{max}} = 441 \text{ nm}$) and compound A ($\lambda_{\text{max}} = 381 \text{ nm}$) measured in acetone or 10% aqueous methanol using a Hewlett-Packard photo-diode array spectrophotometer. Both non-fluorescent compounds were isolated by HPLC, purified using a Sep-Pak C-18 cartridge, passed again through the chromatography column and isolated, dried under N_2 , and dissolved in HPLC-grade solvents before spectrophotometry. Spectra are similar to those of authentic scytonemin (supplied by F. Garcia-Pichel, Max Planck Institute for Marine Microbiology, Bremen). Retention times of compound A (6.1 min) and compound B (5.7 min) were in the range of those of fucoxanthin from diatoms (5.0 min), violaxanthin from higher plants (6.4 min) and the chromatographic dye Sudan II (7.3 min). **b**, Production of UV-radiation-specific pigment (compound A; mean $\pm$ s.e., $n = 6$) on mud substrates in Pipit Lake, Alberta, Canada, exposed to natural levels of UV radiation (+UV, screened with Acrylite OP4), no UV radiation (–UV; Plexiglas UF3), or UV-A radiation alone (–UV-B; Mylar D). UV radiation screens (120 $\times$ 60 cm) were suspended 15 cm over an enclosed 50-cm water column for 30 d. Pigments in the surface 5 mm of lake mud were sampled using standard gravity core procedures²⁹. **c**, Abundance of compound A relative to the sum of indicator carotenoids (alloxanthin, lutein-zeaxanthin, diatoxanthin) as a function of average lake depth and summer DOC concentration. Relative abundance of compound A is greatest in lakes with low integral attenuation of UV radiation. Sedimentary compounds A and B were quantified assuming each had structural and light-extinction characteristics similar to those of carotenoid lutein. Fossil concentrations of compounds A and B were expressed relative to the sum of indicator algal pigments so that the UV radiation index would vary as a function of UV radiation or DOC, independent of algal abundance.

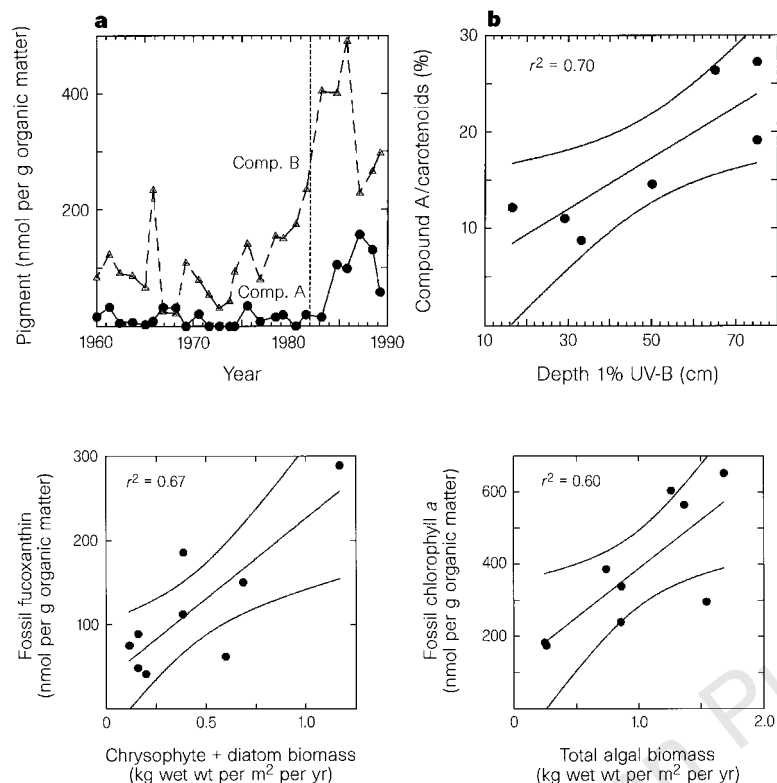


Figure 2 **a**, Concentration of fossil UV-radiation-specific pigments (compounds A and B) as a function of time in Lake 302. These concentrations increased three- to ten-fold following lake acidification (1982–90), loss of DOC and increased penetration of UV radiation. **b**, The relative abundance of compound A was correlated ($P < 0.05$) to the mean depth of UV-B penetration between 1978 and 1987. UV-radiation-specific pigments were not detected in cores from three shallow reference lakes with surface blooms of algae, but low UV radiation transparency.

Figure 3 Concentrations of labile^{8,19} fossil pigments (top, fucoxanthin; bottom, chlorophyll *a*) in Lake 302S were correlated ($P < 0.05$) to the biomass of their respective algal populations between 1978 and 1990. Fucoxanthin is derived mainly from chrysophytes and diatoms^{8,29}, whereas chlorophyll *a* is ubiquitous among algae⁸. Phytoplankton were collected and quantified using consistent, standard techniques^{19,30}. Linear correspondence between algal and fossil abundance demonstrates that stratigraphic patterns of even highly labile fossil pigments in Lake 302S were not artefacts arising from post-depositional degradation, such as can occur in other lakes^{8,19}.

abundance and community composition of primary producers in each lake^{8,19}. Pigments were extracted, isolated, identified and quantified using standard procedures, including high-performance liquid chromatography, HPLC²⁰. Recent calibrations of fossil pigment concentrations against long-term historical records suggest that sedimentary pigments can be as reliable as morphological fossils from algae and can accurately record major changes in algal abundance and community composition¹⁹.

Sediments from many shallow lakes contained two unique non-fluorescent pigments that represent environments of high UV radiation flux (Fig. 1a). Compound A strongly absorbed UV radiation (wavelength of maximum absorbance $\lambda_{\max} = 381$ nm), whereas carotenoid-like compound B had $\lambda_{\max} = 441$ nm. These spectra are similar to the oxidized (compound A) and alkaline (compound B) forms of scytonemin, a pigment produced in the sheaths of certain cyanobacteria on exposure to UV radiation^{9,21}. However, both sedimentary pigments were distinguishable from authentic scytonemin on our HPLC system²⁰. We infer that compounds A and B are biological products of sediment-dwelling algae because intensive sampling in three alpine lakes^{11,13} revealed that the pigments were absent from the water column and rock substrates, yet were abundant in soft sediments. As well, experimental elimination of either UV-B (280–320 nm) or total UV radiation (< 400 nm) using differentially-absorbent aerial screens significantly reduced (repeated measures analysis of variance, $P < 0.05$) production of both compound A and B on the bottom of alpine Pipit Lake^{11,13} relative to mud that received natural irradiances, consistent with the photo-protective function of these pigments (Fig. 1b). Further characterization of compounds A and B are underway.

Pigments specific to the presence of UV radiation were most abundant in shallow lakes (mean depth < 5 m) with low concentrations of UV-radiation-absorbing DOC (< 2 mg l⁻¹; Fig. 1c). In general, relative abundance of compound A was greatest in alpine lakes, the sites with the highest ambient UV radiation^{14,15} and lowest mean concentrations of DOC (0.9 ± 0.7 mg l⁻¹; $n = 25$). Because

compounds A and B were only abundant in lakes with low integral attenuation of UV radiation, we suggest that high relative abundance of these compounds may be a useful marker of an increased depth of penetration of UV radiation in lakes. We tested this hypothesis by analysing historical changes in compounds A and B in a temperate lake known to have undergone an eight-fold increase in penetration by UV radiation following experimental acidification⁴.

Experimental acidification of boreal Lake 302S (49°02' N, 93°42' W) with sulphuric acid reduced lakewater pH from 6.6 to 4.5, reduced DOC from ~ 7 to ~ 1.5 mg l⁻¹, and increased the maximum depth of UV-B penetration eight-fold between 1982 and 1990⁴. We used high-resolution freeze-coring techniques¹⁹ to collect undisturbed sediment cores from the deepest portion of Lake 302S and three local reference lakes in January 1991. Sediments were sectioned in 3-mm intervals and analysed for pigment content using standard procedures¹⁹. Sediment age was estimated using accumulation rates from earlier ²¹⁰Pb-dated cores^{19,22} and, for Lake 302S, from changes in the fossil community composition of scaled chrysophyte algae, reliable indicators of lakewater pH²³. Fossil chrysophyte community composition changed from *Mallomonas duerrschmidtiae* to acidophilic *Synura echinulata* dominance, concomitant with documented acidification in 1982. Inferences of pH change derived from chrysophytes²³ were highly correlated ($r^2 = 0.85$, $P < 0.001$) with measured pH, demonstrating that the fossil chronology in Lake 302S was reliable.

Concentrations of fossil compounds A and B increased three- to ten-fold following the addition of acid to Lake 302S (Fig. 2a). These pigments were not recorded either in sediments from less acidic Lake 302N (3.7–5.0 mg DOC l⁻¹; 1989–90), or in nearby unacidified lakes¹⁹. In particular, compounds A and B were absent from shallow reference lakes 227 and 226, eutrophic sites with extensive cyanobacteria populations¹⁹ but high DOC content (~ 6 mg l⁻¹), suggesting that surface blooms of algae do not produce signals of high transparency to UV radiation.

When expressed as a fraction of indicator carotenoid abundance,

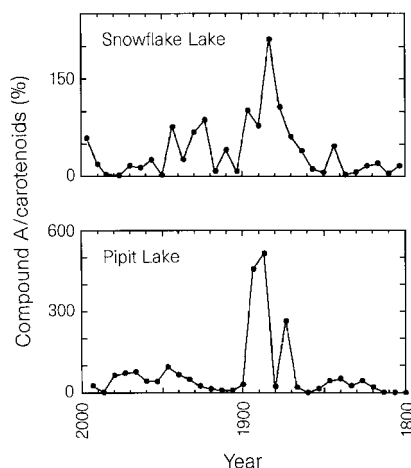


Figure 4 Regional increases in penetration of UV radiation in Snowflake Lake (top) and Pipit Lake (bottom) following droughts in western Canada^{26,28}, indicated by elevated concentration of UV-radiation-specific pigment (compound A) from ~1850 to ~1900. Fossil compound A data are from ref. 11.

concentrations of UV-radiation-absorbing compound A were significantly correlated ($P < 0.05$) with the maximum depth of UV radiation penetration between 1978 and 1987 (Fig. 2b). Although the relationship was less reliable in the uppermost surface sediments (6 mm; 1988–90), these surficial deposits are frequently difficult to interpret because processes that form the fossil record are incomplete⁸. Overall stratigraphic patterns are unlikely to have resulted from post-depositional pigment degradation both because the deep-water oxygen content, and hence preservation environment^{8,19}, has varied little in Lake 302 from 1972 to 1990^{4,24}, and because concentrations of even labile^{8,19} fossil pigments were linearly correlated ($r^2 = 0.60–0.67$, $P < 0.05$) with the abundance of their respective producer populations (Fig. 3). Relative abundances of compounds A and B were lower in Lake 302S than in the most transparent alpine lakes, consistent with greater atmospheric attenuation of UV radiation at low elevations^{14,15}, the relatively great mean depth of Lake 302S (>5 m), and residual reservoirs of UV-radiation-absorbing DOC (> 1.4 mg l⁻¹).

Pigments specific to the presence of UV radiation also recorded regional variations in UV penetration that may periodically harm aquatic organisms in undisturbed lakes (Fig. 4). For example, reanalysis of previously studied cores¹¹ demonstrates that the relative abundance of compounds A and B increased sharply from ~1850 to 1900 in both Snowflake and Pipit lakes (51°36'N, 115°50'W), two similar alpine lakes located within 5 km of each other within Banff National Park, Canada¹¹. Peak abundances of compounds A and B were greater than those exhibited by many present-day mountain lakes (Fig. 1c). This episode of greatly increased penetration of UV radiation occurred at the same time as droughts at low elevations²⁵ and cool temperatures at the tree line in western Alberta²⁶, as inferred from tree-ring analyses. Recent studies of poorly-buffered Precambrian Shield lakes have demonstrated that brief droughts can increase penetration of UV radiation by reducing concentrations of UV-radiation-absorbent DOC^{5,27,28}. Our results suggest that droughts may affect a much wider range of lakes than previously supposed, including undisturbed, well

buffered¹¹ mountain sites and other locations with naturally low concentrations of DOC. Biological damage may be particularly severe in alpine lakes because they receive (and contain) less photo-protective terrestrial DOC (0.9 ± 0.7 mg l⁻¹; $n = 25$) than do subalpine sites (4.7 ± 4.4 mg l⁻¹; $n = 37$).

More than 100,000 North American boreal lakes⁴ and many mountain lakes (Fig. 1) have such low concentrations of DOC (≤ 2 mg DOC l⁻¹) that even slight reductions in DOC following droughts^{4,5,27,28} or acidification^{4,5} can vastly increase penetration of UV radiation. Because this penetration increases as a negative exponential function of DOC concentration^{4,12,17}, drought-induced increases in UV radiation can exceed those anticipated from stratospheric ozone depletion⁴. Analysis of pigments in lake sediments provides a new tool to quantify the magnitude and regional extent of changes in UV-radiation penetration, particularly for lakes which lack historical data on UV radiation. □

Received 9 September 1996; accepted 20 June 1997.

- Bothwell, M. L., Sherbot, D. & Pollock, C. M. Ecosystem response to solar ultraviolet-B radiation: Influence of trophic-level interactions. *Science* **265**, 97–100 (1994).
- Karentz, D. *et al.* Impact of UV-B radiation on pelagic freshwater ecosystems: Report of working group on bacteria and phytoplankton. *Arch. Hydrobiol.* **43**, 31–69 (1994).
- Siebeck, O. *et al.* Impact of UV-B radiation on zooplankton and fish in pelagic freshwater ecosystems. *Arch. Hydrobiol.* **43**, 101–114 (1994).
- Schindler, D. W., Curtis, J. P., Parker, B. R. & Stainton, M. P. Consequences of climate warming and lake acidification for UV-B penetration in North American boreal lakes. *Nature* **379**, 705–708 (1996).
- Yan, N. D., Keller, W., Scully, N. M., Lean, D. R. S. & Dillon, P. Increased UV-B penetration in a lake owing to drought-induced acidification. *Nature* **381**, 141–143 (1996).
- Fritz, S. C., Juggins, S., Battarbee, R. W. & Engstrom, D. R. Reconstruction of past changes in salinity and climate using a diatom-based transfer function. *Nature* **352**, 706–708 (1991).
- Douglas, M. S. V., Smol, J. P. & Blake, W. Jr Marked post-18th century environmental change in high Arctic ecosystems. *Science* **266**, 416–419 (1994).
- Leavitt, P. R. A review of factors that regulate carotenoid and chlorophyll deposition and fossil pigment abundance. *J. Paleolimnol.* **9**, 109–127 (1993).
- Garcia-Pichel, F. & Castenholz, R. W. Characterization and biological implications of scytonemin, a cyanobacterial sheath pigment. *J. Phycol.* **27**, 395–409 (1991).
- Karentz, D., McEuen, F. S., Land, M. C. & Dunlap, W. C. Survey of mycosporine-like amino acid compounds in Antarctic marine organisms: Potential protection from ultraviolet exposure. *Mar. Biol.* **108**, 157–166 (1991).
- Leavitt, P. R., Schindler, D. E., Paul, A. J., Hardie, A. K. & Schindler, D. W. Fossil pigment records of phytoplankton in trout-stocked alpine lakes. *Can. J. Fish. Aquat. Sci.* **51**, 2411–2423 (1994).
- Scully, N. M. & Lean, D. R. S. The attenuation of ultraviolet radiation in temperate lakes. *Arch. Hydrobiol.* **43**, 135–144 (1994).
- Vinebrooke, R. D. & Leavitt, P. R. Effects of ultraviolet radiation on periphyton in an alpine lake. *Limnol. Oceanogr.* **41**, 1035–1040 (1996).
- Blumthaler, M. & Ambach, W. Indication of increasing solar ultraviolet-B radiation flux in alpine regions. *Science* **248**, 206–208 (1990).
- Caldwell, M. M., Robberecht, R. & Billings, W. D. A steep latitudinal gradient of solar ultraviolet-B radiation in the arctic-alpine life zone. *Ecology* **61**, 600–611 (1980).
- Baron, J., McKnight, D. & Denning, A. S. Sources of dissolved and particulate organic material in Loch Vale watershed, Rocky Mountain National Park, Colorado, USA. *Biogeochemistry* **15**, 89–110 (1991).
- Morris, D. P. *et al.* The attenuation of solar UV radiation in lakes and the role of dissolved organic carbon. *Limnol. Oceanogr.* **40**, 1381–1391 (1995).
- Anderson, R. S. Crustacean plankton communities of 340 lakes and ponds in and near the national parks of the Canadian Rocky Mountains. *J. Fish. Res. Board Can.* **31**, 855–869 (1974).
- Leavitt, P. R. & Findlay, D. L. Comparison of fossil pigments with 20 years of phytoplankton data from eutrophic Lake 227, Experimental Lakes Area, Ontario. *Can. J. Fish. Aquat. Sci.* **51**, 2286–2299 (1994).
- Mantoura, R. F. C. & Llewellyn, C. A. The rapid determination of algal chlorophyll and carotenoid pigments and their breakdown products in natural waters by reverse phase high-performance liquid chromatography. *Anal. Chim. Acta* **151**, 297–314 (1983).
- Proteau, P. J., Gerwick, W. H., Garcia-Pichel, F. & Castenholz, R. W. The structure of scytonemin, an ultraviolet sunscreen pigment from the sheaths of cyanobacteria. *Experimentia* **49**, 825–829 (1993).
- Anderson, R. F., Schiff, S. L. & Hesslein, R. H. Determining sediment accumulation and mixing rates using ²¹⁰Pb, ¹³⁷Cs, and other tracers: Problems due to postdepositional mobility or coring artifacts. *Can. J. Fish. Aquat. Sci.* (suppl. 1) **44**, 231–250 (1987).
- Dixit, S. S., Dixit, A. S. & Smol, J. P. Lake acidification recovery can be monitored using chrysophycean microfossils. *Can. J. Fish. Aquat. Sci.* **46**, 1309–1312 (1989).
- Schindler, D. W., Rusczyński, T. & Fee, E. J. Hypolimnetic injection of nutrient effluents as a method for reducing eutrophication. *Can. J. Fish. Aquat. Sci.* **37**, 320–327 (1980).
- Case, R. A. & MacDonald, G. M. A dendroclimatic reconstruction of annual precipitation on the western Canadian prairies since A.D. 1505 from *Pinus flexilis* James. *Quat. Res.* **44**, 267–275 (1995).
- Luckman, B. H. Mountain areas and global change: A view from the Canadian Rockies. *Mount. Res. Dev.* **10**, 183–195 (1990).
- Schindler, D. W. *et al.* Climate-induced changes in the dissolved organic carbon budgets of boreal lakes. *Biogeochemistry* **36**, 9–28 (1997).
- Schindler, D. W. *et al.* The effects of climatic warming on the properties of boreal lakes and streams at the Experimental Lakes Area, northwestern Ontario. *Limnol. Oceanogr.* **41**, 1004–1017 (1996).
- Goodwin, T. W. *The Biochemistry of the Carotenoids* Vol. 1 Plants (Chapman & Hall, London, 1980).
- Findlay, D. L. & Kasian, S. E. M. Phytoplankton communities of lakes experimentally acidified with sulfuric and nitric acids. *Can. J. Fish. Aquat. Sci.* **47**, 1378–1386 (1990).

Acknowledgements. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Correspondence should be addressed to P.L. (e-mail: Leavitt@leroy.cc.uregina.ca).

Predicted reduction in basal melt rates of an Antarctic ice shelf in a warmer climate

K. W. Nicholls

British Antarctic Survey, Natural Environment Research Council,
Cambridge CB3 0ET, UK

Floating ice shelves are vulnerable to climate change at both their upper and lower surfaces. The extent to which the apparently air-temperature-related retreat of some northerly Antarctic Peninsula ice shelves¹ presages the demise of their much larger, more southerly, counterparts is not known, but air-temperature effects are unlikely to be important in the near future. Oceanographic measurements from beneath the most massive of these southerly ice shelves—the Filchner–Ronne Ice Shelf^{2–4}—have confirmed that dense sea water resulting from sea-ice formation north of the ice shelf flows into the sub-ice-shelf cavity. This relatively warm so-called High Salinity Shelf Water (HSSW) is responsible for the net melting at the ice shelf's base. Here I present temperature measurements, from the same sub-ice-shelf cavity, which show a strong seasonality in the inflow of HSSW. This seasonality results from intense wintertime production of sea ice, and I argue that the seasonal springtime warming can be used as an analogue for climate warming. For the present mode of oceanographic circulation, the implication is that warmer winters (a climate warming, leading to lower rates of sea-ice formation, would cause a reduction in the flux of HSSW beneath the ice shelf. The resultant cooling in the sub-ice cavity would lead, in turn, to a reduction in the total melting at the ice shelf's base. A moderate warming of the climate could thus lead to a basal thickening of the Filchner–Ronne Ice Shelf, perhaps increasing its longevity.

The Filchner–Ronne Ice Shelf lies in the southern Weddell Sea (Fig. 1), where mean annual air temperatures are more than 10 °C lower than the critical value, –5 °C, thought to be the limit of ice-shelf viability¹. Even at the present high rates of warming recorded on the Antarctic Peninsula to the north, 2.5 °C over the past 50 years, it would take 200 years for this safety margin of 10 °C to be eroded. A pertinent question is therefore: how, if at all, will the larger ice shelves respond in the near future?

The most likely threat to the larger, more southerly ice shelves derives from their intimate thermal contact with the underlying ocean. The basic features of the circulation beneath the Ross Ice Shelf were described by MacAyeal⁵, and it seems that a similar mode exists beneath the Filchner–Ronne Ice Shelf^{2–4}. HSSW formed north of the ice front (the seaward limit of the ice shelf) drains down the sea-floor slope towards the grounding line, where the inland ice goes afloat to form the ice shelf. There it comes into contact with the base of the ice shelf at depths up to 1,500 m below sea level. At such depths the pressure depresses the freezing point of sea water by more than 1 °C below the temperature of HSSW (–1.9 °C), and strong basal melting ensues. The HSSW becomes mixed with melt water and leaves the cavity as Ice Shelf Water, a cold water mass that contributes to the production of Antarctic Bottom Water^{6–8}. The formation of Ice Shelf Water, and its contribution to the production of Antarctic Bottom Water, establishes the Filchner–Ronne Ice Shelf as a significant element in the world climate system. It has been argued⁵ that a warming climate will have little direct effect on melt rates beneath ice shelves whose underlying ocean cavities are ventilated by HSSW because the temperature of HSSW is fixed at the surface freezing point of sea water (–1.9 °C). Here I present new data which imply that, providing the large-scale oceanographic regime of the southern Weddell Sea remains

unchanged, a warming of the climate will (if anything) reduce the basal melt rate of the Filchner–Ronne Ice Shelf.

In 1990, the British Antarctic Survey began a programme to make direct oceanographic measurements beneath the Ronne Ice Shelf^{2–4}, the part of the Filchner–Ronne system lying to the west of Berkner Island. A hot-water drill was used to create access holes at three sites (Fig. 1), hundreds of kilometres from the ice front, allowing the use of an oceanographic probe to obtain profiles of temperature and salinity from beneath the ice shelf. The most recently occupied site, site 3, was drilled during January 1996⁴. At this site, 825 m of ice overlies a 485-m-deep water column. In addition to the measurement of temperature and salinity profiles, work at site 3 included the deployment of an instrument string with current meters and conductivity/temperature instrument packages. The location of site 3 was selected in an attempt to monitor the inflow of HSSW that had previously been observed at site 2. The broken line in Fig. 1 shows the probable path of the HSSW under study⁴.

The top trace in Fig. 2 shows a year-long temperature record obtained from an instrument package located 90 m above the sea floor at site 3. The record has been filtered to remove variability at periods of less than about a month. The solid line represents the available data, the broken line merely duplicates the data for future years on the assumption that the variation is seasonal. The depth of the instrument package was chosen to place it in the middle of a mixed layer at the bottom of the water column. A current meter at the same depth indicated that the water in the mixed layer was flowing deeper into the cavity at a speed of ~4.6 cm s^{–1}. At –1.96 °C, the water is cooler than pure HSSW (–1.90 °C), and it has presumably mixed with colder water on its journey from the ice front to

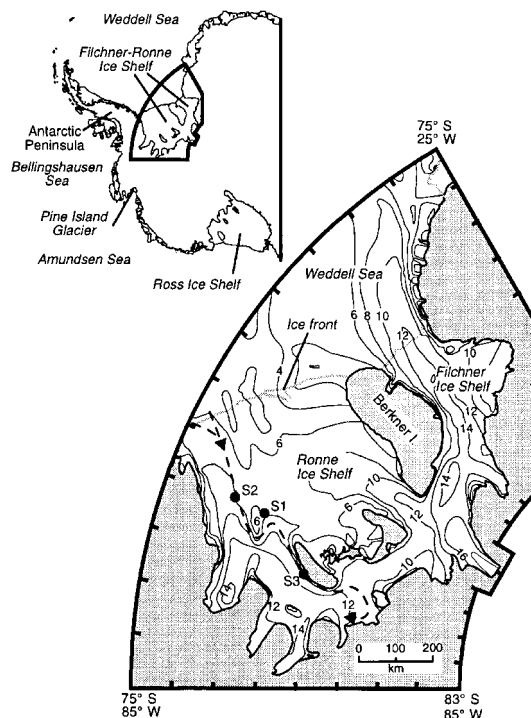


Figure 1 Top left, map showing the position of the Filchner–Ronne Ice Shelf in the southern Weddell Sea, and (main figure) an enlargement of the region of the ice shelf itself and part of the adjacent continental shelf. Shading indicates areas of grounded ice. The contours give the bathymetry in hundreds of metres below sea level^{15,16}, and the sites are marked S1, S2 and S3. The High Salinity Shelf Water (HSSW) flows as a gravity current and is therefore expected to follow the bathymetry. The broken line shows the presumed route of the HSSW.

site 3 (ref. 4). The variability in the record is interpreted as resulting from a seasonal thinning of the lower mixed layer, causing the instrument package to emerge into the colder Ice Shelf Water above. A similar effect was observed at site 2 (ref. 9). A 22-month water-temperature record obtained from 33 m above the sea floor at site 2 is shown in the middle trace in Fig. 2. The record has been filtered in the same way as the site-3 record, and again shows a seasonal thinning of a lower layer of HSSW.

During the summer months, sea-ice formation in the open shore lead north of the ice front ceases, and the production of HSSW presumably reduces dramatically. The cessation of ice production is evident from summer oceanographic cruises in the shore lead which have revealed a layer of warm water at the sea surface^{10,11}. Between 1989 and 1993 an air-temperature record was obtained from an automatic weather station near the Ronne Ice Front, and is shown on an inverted scale in the bottom trace in Fig. 2. The record was filtered in the same way as the water-temperature records, and indicates a summer about 4 months in length, lasting from mid-October to mid-February. Between the onset of summer and the onset of the cooling at sites 2 and 3 (thinning of the HSSW layer) there is a delay that is caused by the time taken for the dense water to drain down into the deeper sub-ice-shelf cavity. The measured delay will, of course, depend on how close to the top of the HSSW layer the thermistor was initially deployed; and at site 3 it is quite possible that the response to the summer warming is delayed by longer than a year.

The observation of seasonality in the thickness of the HSSW flow at sites 2 and 3 indicates that the volume flux of HSSW into the sub-ice-shelf cavity responds to the seasonal warming of the air north of the ice front. Summer conditions do not increase the temperature of the inflow into this cavity: HSSW is the only water mass dense enough to sink beneath the ice shelf, and its temperature is fixed at the surface freezing point. Summer conditions evidently do, though, result in a reduction in the volume of HSSW flowing into the sub-ice-shelf cavity, and therefore the flux of heat. Less intense wintertime sea-ice production resulting from climate warming would have a similar effect, reducing the HSSW flow, and therefore

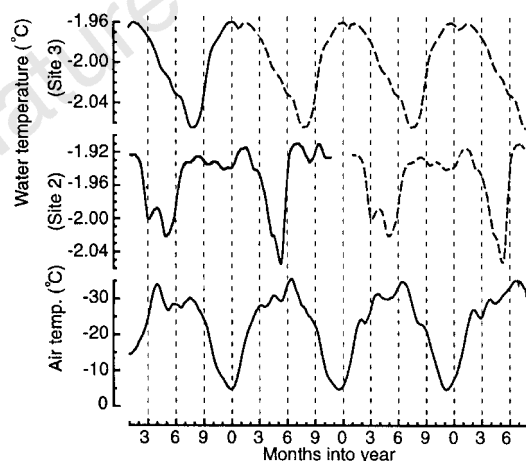


Figure 2 Top trace, the temperature record at site 3 taken 90 m above the sea floor, obtained during 1996. Middle trace, temperature record at site 2, collected during 1992 and 1993. Bottom trace, air-temperature record obtained between 1989 and 1992 from near the Ronne Ice Front. Note that none of the time series are concurrent, and that the broken portions of the top and middle traces are repetitions of the observations (the solid lines) to help show the timings of the variations. All three records have been low-pass-filtered to remove variability on timescales shorter than about a month. On the horizontal axis, months 0, 3, 6 and 9 represent January, March, June and September, respectively.

the heat flux, beneath the ice shelf. In the first instance this will reduce the rate of melting from the base of the Ronne Ice Shelf, particularly in the region of the grounding lines where the HSSW makes the first contact with the ice⁷. The oceanographic outcome would be a reduction in the total flux of Ice Shelf Water from beneath the ice shelf. For the ice shelf itself, the effect would be to allow the total ice volume to increase.

Sea-ice formation rates depend not only on the air temperature, but also on the wind speed, and on the area of open water exposed to the atmosphere which, again, is primarily controlled by wind speed and direction. The assumption that a warmer climate leads to lower wintertime sea-ice production therefore depends on these other factors either remaining constant or reducing. Coupled general circulation models are at present not thought reliable for predicting regional-scale responses to global climate forcing. However, data from the Hadley Centre climate model have been analysed to find the expected response in the region of the Ronne Ice Shelf to a global increase in greenhouse gases (experiment SUL¹²). The model predictions of a 3 °C regional warming over 100 years, and a 10% reduction in wind speeds, support the assertion that a warmer climate will result in reduced rates of sea-ice formation north of the Ronne Ice Front. Results from the model of Grumbine¹³ indicate that HSSW formation rates would decrease as a consequence. The effect on the basal melt rate of the ice shelf would be modest: if the rate of sea-ice production is proportional to the difference between the wintertime temperatures of the sea surface (−1.9 °C) and the air (−30 °C, see Fig. 2) then a 10% decrease in that difference (∼3 °C) would be expected to reduce overall basal melt rates by ∼10%.

If the climate were to warm even further, other effects are likely to dominate. In the Bellingshausen and Amundsen seas (Fig. 1) the warm Circumpolar Deep Water surrounding Antarctica⁸ is able to flood onto the continental shelf, and this is the probable reason for the lack of large ice shelves in that sector. Basal melt rates for ice shelves that do lie in the Amundsen/Bellingshausen Sea sector are one or two orders of magnitude greater than for the Filchner–Ronne Ice Shelf, an average value of 12 m yr^{−1} being reported for the floating portion of Pine Island Glacier in the Amundsen Sea¹⁴ (Fig. 1). Weddell Deep Water, the slightly cooler Weddell Sea derivative of Circumpolar Deep Water, is absent from the Weddell Sea continental shelf. It is unclear how this aspect of the Weddell Sea regional oceanography will respond to an extreme climate warming. Clearly, though, a large-scale incursion of Weddell Deep Water is the primary danger to the Filchner–Ronne Ice Shelf. At present a tongue of Modified Weddell Deep Water penetrates as far south as the Ronne Ice Front¹¹ but is unlikely to have much effect on the basal melt rate because of the relatively dense HSSW that prevents it from flowing deep under the ice shelf. A large enough decrease in HSSW production might allow Modified Weddell Deep Water to penetrate further into the sub-ice-shelf cavity, causing melt-rates to rise. Whether this is likely in the foreseeable future is not yet known.

From the arguments presented here, a seasonal warming results in a corresponding reduction in the melting at the base of the Filchner–Ronne Ice Shelf, an effect expected to be strongest near the deep grounding lines. To the extent that this is an analogue for a warming climate, it suggests that, for as long as HSSW dominates the oceanographic conditions over the continental shelf, the response of the ice shelf to a warming of the climate will be for it to thicken, reinforcing rather than threatening its longevity. □

Received 7 November 1996; accepted 17 June 1997.

1. Vaughan, D. G. & Doake, C. S. M. Recent atmospheric warming and retreat of ice shelves on the Antarctic Peninsula. *Nature* **379**, 328–331 (1996).
2. Nicholls, K. W., Makinson, K. & Robinson, A. V. Ocean circulation beneath the Ronne ice shelf. *Nature* **354**, 221–223 (1991).
3. Robinson, A., Makinson, K. & Nicholls, K. The oceanic environment beneath the north-west Ronne Ice Shelf, Antarctica. *Ann. Glaciol.* **20**, 386–390 (1994).
4. Nicholls, K. W., Makinson, K. & Johnson, M. R. New oceanographic data from beneath Ronne Ice Shelf, Antarctica. *Geophys. Res. Lett.* **24**, 167–170 (1997).

5. MacAyeal, D. R. Thermohaline circulation below the Ross Ice Shelf: A consequence of tidally induced vertical mixing and basal melting. *J. Geophys. Res.* **89**, 597–606 (1984).
6. Jacobs, S. S., Amos, A. F. & Bruchhausen, P. M. Ross Sea oceanography and Antarctic Bottom Water formation. *Deep-Sea Res.* **17**, 935–962 (1970).
7. Jacobs, S. S., Fairbanks, R. G. & Horibe, Y. in *Oceanology of the Antarctic Continental Shelf* (ed. Jacobs, S. S.) 59–85 (Antarctic Res. Ser. 43, Am. Geophys. Un., Washington DC, 1985).
8. Foldvik, A. & Gammelsrød, T. Notes on Southern Ocean hydrography, sea-ice and bottom water formation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **67**, 3–17 (1988).
9. Nicholls, K. W. Temperature variability beneath Ronne Ice Shelf, Antarctica, from thermistor cables. *J. Geophys. Res.* **101**, 1199–1210 (1996).
10. Foldvik, A., Gammelsrød, T. & Tørresen, T. in *Oceanology of the Antarctic Continental Shelf* (ed. Jacobs, S. S.) 5–20 (Antarctic Res. Ser. 43, Am. Geophys. Un., Washington DC, 1985).
11. Gammelsrød, T. et al. in *The Polar Oceans and their Role in Shaping the Global Environment* (eds Johannessen, O. M., Muench, R. D. & Overland, J. E.) 159–176 (Geophys. Monogr. 84, Am. Geophys. Un., Washington DC, 1994).
12. Mitchell, J. F. B., Johns, T. C., Gregory, J. M. & Tett, S. F. B. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature* **376**, 501–504 (1995).
13. Grumbine, R. W. A model of the formation of High-Salinity Shelf Water on polar continental shelves. *J. Geophys. Res.* **96**, 22049–22062 (1991).
14. Jacobs, S. S., Hellmer, H. & Jenkins, A. Antarctic Ice Sheet melting in the Southeast Pacific. *Geophys. Res. Lett.* **23**, 957–960 (1996).
15. Johnson, M. R. & Smith, A. M. Seabed topography under the southern and western Ronne Ice Shelf, derived from seismic surveys. *Antarct. Sci.* **9**, 201–208 (1997).
16. Vaughan, D. G. et al. *Map of subglacial and seabed topography 1:2000000 Filchner-Ronne-Schelfeis, Antarktis* (Institut für Angewandte Geodäsie, Frankfurt am Main, Germany, 1994).

Acknowledgements. I thank all those who have helped with the fieldwork, and W. M. Connolly for analysis of model data.

Correspondence should be addressed to K.W.N. (e-mail: k.nicholls@bas.ac.uk).

Laboratory simulations of sustained volcanic eruptions

H. M. Mader*, E. E. Brodsky†, D. Howard‡ & B. Sturtevant‡

* Department of Geology, University of Bristol, Wills Memorial Building, Queens Road, Bristol, BS8 1RJ, UK

† Seismological Laboratory, ‡ Graduate Aeronautical Laboratories, California Institute of Technology, Pasadena, California 91125, USA

Many violent eruptions are driven by rapid exsolution of dissolved volatiles within liquid magma. The accelerating two-phase mixture emerges at the vent as a sustained quasi-steady discharge lasting for periods from hours to days (refs 1, 2). The initial growth in discharge rate commonly observed^{2–4}, subsequent fluctuations^{5,6} and discrete pulses and shocks⁷ remain largely unexplained. We have simulated volcanic conduit flows by producing sustained, quasi-steady explosions in a liquid undergoing rapid exsolution of a gas. This was done by rapidly decompressing large volumes of CO₂-saturated water. The results reveal fluctuations in discharge rate that reflect heterogeneities in the two-phase mixture that form spontaneously as a consequence of the size and geometry of the experimental system. An initial transient with a growing discharge rate is observed in experiments in which material is erupted from a spherical flask up a narrow neck that mimics the magma-chamber/conduit assembly of volcanic systems. The fragmentation region propagates down the neck during the initial transient until it reaches a stable position at the top of the flask, at which point a quasi-steady discharge ensues.

The experiments were carried out (Fig. 1) using a shock-tube facility at the California Institute of Technology, previously described by Mader et al.⁸. Until now, exsolution experiments using this apparatus have been restricted to the study of discrete, transient explosions (typical durations of a few milliseconds) using small sample volumes (100 ml or less)⁸. Other shock-tube experiments have been similar in this respect^{8–11}, although quasi-steady flows of short duration (~20 ms) have been generated by explosive vaporization^{12,13}. In the experiments reported here, much larger sample volumes, between 0.5 and 1.8 l, and two different test cells are used. The first arrangement (Fig. 1a) uses a cylindrical test cell with a constriction at the top, which has the effect of slowing down

the eruption and thereby increasing its duration. Experiments in which the constriction is omitted have a shorter duration but show qualitatively the same behaviour. In the second arrangement (Fig. 1b), material is erupted from a large sphere up a narrow neck.

The problem of scale is central to understanding the laboratory eruptions. A true simulation involves reproducing the flow dynamics of a large-scale natural system within a small-scale model by matching the ratios of important forces. But we do not know which of the many force ratios within the bulk flow control the regime of these high-density two-phase flows. Similarity can be achieved on the scale of individual particles, but this will not necessarily lead to similar bulk flow behaviour, because the overall flow is dominated by particle–particle interactions. Thus, the approach taken is to replicate as many of the dynamical parameters of the system is possible, while providing a wide range of spatial scales, so that the long-range particle–particle interactions can act. Velocities and accelerations in our experiments (locally up to 20 m s⁻¹ and 100 g) are similar to those found in nature^{1,2}. Gravitational effects (1 g) do not control the dynamics of flows that experience such large accelerations and are therefore not modelled. Viscous forces are also not scaled. But numerical calculations indicate that, even under conditions of explosive eruptions, diffusive bubble growth is not retarded by viscous effects unless magma viscosities exceed 10⁸ Pa s (refs 14, 15); this condition is not generally reached during fragmentation because magmas are not fully degassed.

The times involved in the laboratory experiments are necessarily scaled down compared to nature by the same factor as the physical dimensions, given similar velocities. This scaling applies also to the duration that may be considered to represent a sustained eruption. Typical eruption durations in the laboratory are between 1,500 and 3,000 ms, about 100 times longer than in previous experiments⁸, and about 50 times the particle transit time up the neck (30–50 ms). This is similar to the ratio of eruption duration (~1 d) and estimated particle transit times between magma chamber and vent (~1,000 s) in typical plinian events⁵. Thus, our laboratory explosions are sustained events that are similar in this respect to natural eruptions.

The physical process driving the explosions is the same as in earlier experiments and so the early behaviour is as previously reported⁸; there is a single nucleation event and bubbles grow in unison. Early in its development, the experiment reaches a state typical of the inception of natural eruptions due to an unloading event. Increases of spatial and temporal scales (sample volumes up to 20 times larger and eruptions durations 100 times longer) provide the opportunity to observe in detail the subsequent flow evolution. All the flows develop pronounced heterogeneities, as shown in Figs 2 and 3. Discrete pulses of finely vesicular material form repeatedly on a scale initially related to the test-cell diameter. These pulses combine, owing to their different velocities, leading to growth in the overall pulse-length with height.

In the flask experiments (Figs 2 and 4), quasi-steady flow conditions are reached after an initial transient, which is complete by $t \approx 400$ ms after diaphragm rupture. During the transient, the discharge rate grows (Fig. 4) and the fragmentation region propagates down the neck until it reaches a stable position at the exit from the sphere (Fig. 2). Thereafter, the flow is characterized by a virtually static foam that fills the sphere and a highly turbulent, heterogeneous and fragmented flow in the neck. This situation persists for ~1,300 ms. The average bulk discharge rate is roughly constant once the initial transient is complete, but small-scale heterogeneities within the flow lead to the formation of discrete pulses (Fig. 2).

These results have a number of important implications for our understanding of explosive volcanic eruption dynamics.

First, the spontaneous and sudden breakup into spatially and temporally highly heterogeneous flows is a result of the increase in scale of the experiments; a larger range of scales is available and the particle–particle interactions have more time and space to evolve.

Further scaling-up to natural dimensions will tend to increase the degree of heterogeneity. In particular, heterogeneities on larger scales, relative to conduit dimensions, should be expected as even longer-range fluid interactions come into play. Thus, we can conclude that volcanic eruptions are inherently heterogeneous because of their scale.

Second, the common observation of discharge rates growing with time during the initial phase of a volcanic eruption has been previously explained in terms of steady erosion of the conduit²⁻⁴. However, this behaviour is closely replicated in our experiments using the flask arrangement, which suggests that independent of any concomitant conduit erosion, a discharge rate increase could be induced by geometrical effects associated with the magma chamber feeding a narrow conduit—a feature of explosive eruption dynamics hitherto ignored by all models of explosive volcanic eruptions¹⁴⁻¹⁸.

The experimental observations can be explained by considering the pressure evolution in the sphere. After diaphragm rupture, a rarefaction wave propagates down into the sphere at the speed of sound in bubble-free water, that is, effectively instantaneously

(<0.5 ms), and bubbles are nucleated. But bubble growth is limited by the rate of gas diffusion into the bubbles and the rate at which material can escape from the sphere, both of which are governed by the pressure in the sphere and bubbles. Quasi-steady flow conditions occur when a pressure balance is reached between gas diffusion into the bubbles and flow out of the sphere. This condition is stable until the CO₂ is depleted and the eruption exhausted.

The flask arrangement is a model for eruptions in which significant bubble nucleation and growth takes place in the magma chamber. This is likely to occur in eruptions that lead to extensive evacuations of the magma chamber. In such eruptions, the discharge rate is necessarily greater than the rate of influx of magma into the chamber at depth. The unloading of the magma chamber will be progressive, rather than instantaneous as in our laboratory system. Nevertheless, a steady-state discharge requires that a similar pressure balance between diffusion into bubbles and flow out of the chamber is attained.

Third, after the initial transient, the fragmentation region remains at the entrance to the neck of the flask. This is because the colliding fluid streams here provide a source of instability within

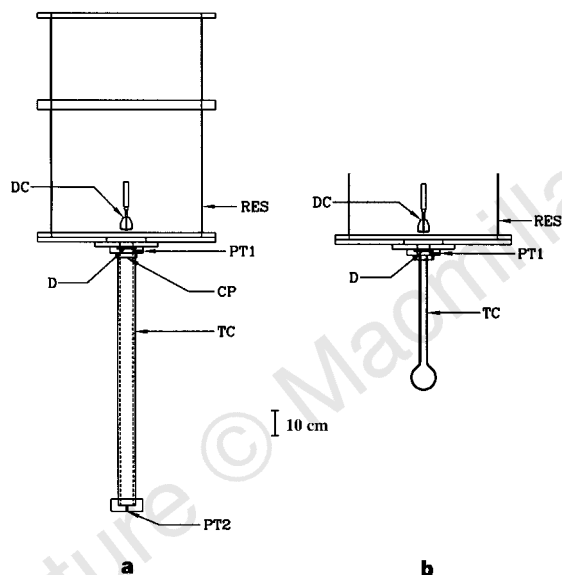


Figure 1 Experimental apparatus. Water that has been saturated under pressure with CO₂ is fed into the high-pressure pyrex test cell TC (pressure $\leq 1,000$ kPa) which is separated from the low-pressure reservoir RES (pressure ~ 4 kPa, volume 260 l) by an aluminium diaphragm D. Sudden decompression and supersaturation is achieved by puncturing the diaphragm with a solenoid-driven cutter DC. The supersaturation causes CO₂ bubbles to nucleate and grow progressively resulting in the expanding two-phase flow which erupts into the reservoir. Pressures are measured at transducers PT1 and PT2. The flows are recorded on video and high-speed motion photography at 1,000 frames per second. **a**, Constricted tube arrangement. The test cell consists of a long cylindrical tube (length 100 cm, diameter 5 cm) with a metal constriction plate CP mounted at the top. The constriction is provided by circular holes of varying diameters (2.54 cm, 1.27 cm and 0.64 cm) drilled centrally in the metal plate. Fill-depths of ~ 50 cm and ~ 90 cm (liquid volumes of between ~ 1 and ~ 1.8 l) and pressures of 750 to 1,040 kPa are used. **b**, Flask arrangement. The test cell consists of a sphere (volume ~ 0.5 l, diameter 10 cm) that grades smoothly into a long cylindrical neck (length 45 cm, diameter 2.7 cm). Fill-depths up to 2 to 10 cm up the neck and a pressure of 760 kPa are used.

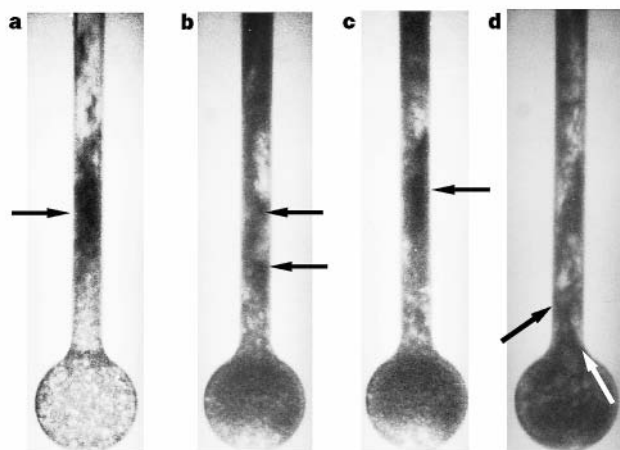


Figure 2 Photographs of results using the flask apparatus (Fig. 1b). The timings of important events are as follows: $t = 0$, diaphragm rupture; $t = 9$ ms, nucleation; $t = 14$ ms, start of flow front motion; fragmentation region propagates down neck at an estimated average speed of the order of 1 m s^{-1} , reaching base of neck at $t \approx 250$ ms; eruption is exhausted at $\sim 1,500$ ms. **a**, Flow at $t = 49$ ms, showing a well-defined fragmentation region a height $h = 8$ cm above the neck entrance (arrowed). The fragmentation region separates the uniformly bubbled liquid below from the fragmented spray above. Flows at $t = 258$ ms (**b**), $t = 268$ ms (**c**) and $t = 1,282$ ms (**d**) have the fragmentation region at the base of the neck separating the homogeneous and virtually static foam in the sphere from the highly heterogeneous and turbulent flow in the neck. The instability at the neck entrance is caused by converging streams of liquid in the sphere, coupled with the high accelerations experienced by the individual fluid particles, leading to large-scale heterogeneities and fragmentation of the liquid into small volumes of finely bubbled (that is, dark) liquid. The finely bubbled regions have very high surface-to-volume ratios and so diffusion in these regions is rapid, resulting in higher accelerations and velocities. The fragments start to gain speed and sweep through the neck, collecting material as they go. The process of slug growth is illustrated in **b** and **c**. In **b**, two diffuse slugs (arrowed) of finely bubbled liquid are visible separated by a region of clear fluid. In **c**, the two slugs have merged to form one dense slug (arrowed). The process of the initial generation of the slugs at the neck entrance is shown in **d**. Converging streams of liquid at the entrance to the neck cause oscillatory behaviour (arrowed) on the scale of the neck diameter.

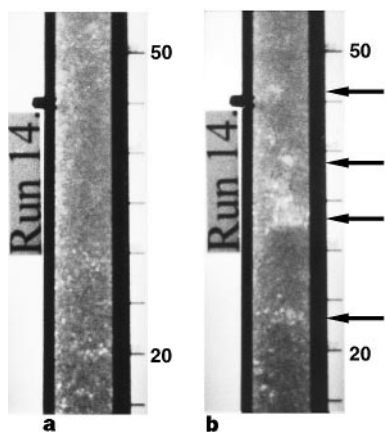


Figure 3 Photographs of results from the constricted-tube apparatus (Fig. 1a). Run conditions: fill depth 89 cm, constriction diameter 1.27 cm, test-cell pressure 737 kPa, reservoir pressure 4 kPa. The bubbles nucleate uniformly throughout the flow in a single event and grow in unison, resulting in the initially homogeneous flow shown in **a** for $t \approx 1.2$ s. A significant breakdown event then occurs leading to the flow shown in **b** 100 ms later, in which finely bubbled slugs of liquid are separated by regions of relatively bubble-free liquid (arrowed). These slugs behave in a qualitatively similar way to those seen in the flask experiments and described in Fig. 2, and are a direct results of the scale of these experiments. The eruption continues for $\sim 3,000$ ms. The numbers on the right of each panel are in centimetres.

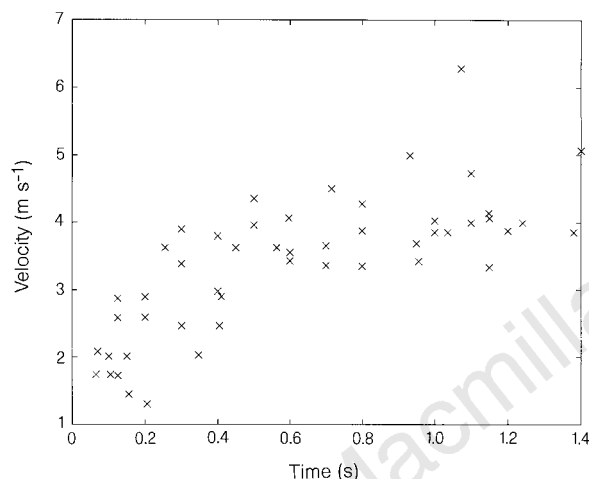


Figure 4 Velocity at the neck entrance of the flask (Fig. 1b) as a function of time. The motion of individual blobs of fluid 1–3 cm from the neck entrance were tracked over 5 ms. The graph shows that the velocity at the neck entrance increases monotonically over the first 400 ± 100 ms of the run until a velocity of ~ 4 m s $^{-1}$ is reached; this remains roughly constant until the eruption is exhausted. The change from increasing velocity to constant velocity at $t \approx 400$ ms corresponds to a break in the slope of the pressure trace measured at the top of the neck. The scatter of the data points is large because (1) the flow is turbulent and so there are real variations in velocity and (2) because the fluid blobs chosen were at variable heights and tended to change shape. The velocities measured at the neck entrance are small, and the slope over the first 400 ms is also small. This slope describes the eulerian acceleration, which is the change in the velocity at a point in the flow as different fluid particles move through it. It is small compared to the lagrangian acceleration, which is the acceleration experienced by a specific fluid particle as it moves past this point. The particles have negligible forward motion in the sphere and then reach a velocity of ~ 4 m s $^{-1}$ over a distance of ~ 2 cm in the neck which corresponds to an average lagrangian acceleration of ~ 200 m s $^{-2}$ or $20g$. The instantaneous accelerations are likely to be several times greater than this.

the flow and also because this is where individual fluid particles experience the greatest accelerations. These conditions are at no time provided within the body of the sphere, no matter how expanded the flow becomes, and so the fragmentation region does not propagate down into it. The acceleration at the neck entrance is a function of the ratio of the typical sizes of neck and sphere. This ratio is likely to be larger in actual volcanic eruptions by at least an order of magnitude compared to our experimental apparatus. Thus, the natural accelerations will be huge as the material enters the conduit making this the most likely location for the fragmentation region.

The second and third implications (above) together provide a simple method for reaching quantitative conclusions about the flow dynamics occurring at depth during a sustained, plinian eruption. The start-up period, with growing magma discharge rate, identifies the time during which the fragmentation region is propagating down the conduit. Therefore, if the conduit length is known, then the maximum average rate of propagation of the fragmentation region can be estimated (for example, 0.6 m s $^{-1}$ for 18 May 1980 Mt St Helens eruption with conduit length ~ 7 km (ref. 19) and start-up period ~ 3 h (ref. 5); the experimental value was ~ 1 m s $^{-1}$ —see Fig. 2). The position of the fragmentation region is defined as being at the top of the magma chamber for quasi-steady flow conditions, separating the static foam in the magma chamber from the turbulent and fragmented flow in the conduit. □

Received 24 January; accepted 10 June 1997.

1. Walker, G. P. L. Plinian eruptions and their products. *Bull. Volcanol.* **44**, 223–240 (1981).
2. Carey, S. & Sigurdsson, H. The intensity of plinian eruptions. *Bull. Volcanol.* **51**, 28–40 (1989).

3. Carey, S. & Sigurdsson, H. Temporal variations in column height and magma discharge rate during the 79 AD eruption of Vesuvius. *Bull. Geol. Soc. Am.* **99**, 303–314 (1987).
4. Wilson, L. Relationships between pressure, volatile content and ejecta velocity in three types of volcanic explosion. *J. Volcanol. Geotherm. Res.* **8**, 297–313 (1980).
5. Carey, S. *et al.* Variations in column height and magma discharge during the May 18 1980 eruption of Mount St Helens. *J. Volcanol. Geotherm. Res.* **43**, 99–112 (1990).
6. Gardner, J. E., Sigurdsson, H. & Carey, S. N. Eruption dynamics and magma withdrawal during the plinian phase of the Bishop tuff eruptions, Long Valley caldera. *J. Geophys. Res.* **96**, 8097–8111 (1991).
7. Sparks, R. S. J. & Wilson, L. Explosive volcanic eruptions. 5. Observations of plume dynamics during the 1979 Soufriere eruption, St Vincent. *Geophys. J. R. Astron. Soc.* **69**, 551–570 (1982).
8. Mader, H. M. *et al.* Experimental simulations of explosive degassing of magma. *Nature* **372**, 85–88 (1994).
9. Mader, H. M. *et al.* Dynamics of explosive degassing of magma: Observations of fragmenting two-phase flows. *J. Geophys. Res.* **101**, 5547–5560 (1996).
10. Anilkumar, A. V., Sparks, R. S. J. & Sturtevant, B. Geological implications and applications of high velocity two-phase flow experiments. *J. Volcanol. Geotherm. Res.* **56**, 145–160 (1993).
11. Phillips, J. C. *et al.* Gum rosin-acetone system as an analogue to the degassing behaviour of hydrated magmas. *Bull. Volcanol.* **57**, 263–268 (1995).
12. Hill, L. G. & Sturtevant, B. in *Adiabatic Waves in Liquid-Vapor Systems* (eds Meier, G. E. A. & Thompson, P. A.) 25–37 (Springer, Berlin, 1990).
13. Sugioka, I. & Bursik, M. Explosive fragmentation of erupting magma. *Nature* **373**, 689–692 (1995).
14. Dobran, F. Nonequilibrium flow in volcanic conduits and application to the eruptions of Mt St Helens on May 18, 1980 and Vesuvius in AD 79. *J. Volcanol. Geotherm. Res.* **49**, 285–311 (1992).
15. Barclay, J., Riley, D. S. & Sparks, R. S. J. Analytical models for bubble growth during decompression of high viscosity magma. *Bull. Volcanol.* **57**, 422–431 (1995).
16. Papale, P. & Dobran, F. Magma flow along the volcanic conduit during the plinian and pyroclastic flow phases of the May 18, 1980, Mount St Helens eruption. *J. Geophys. Res.* **99**, 4355–4373 (1994).
17. Sparks, R. S. J. The dynamics of bubble formation and growth in magmas: a review and analysis. *J. Volcanol. Geotherm. Res.* **3**, 1–37 (1978).
18. Wilson, L., Sparks, R. S. J. & Walker, G. P. L. Explosive volcanic eruptions. IV. The control of magma properties and conduit geometry on eruption column behaviour. *Geophys. J. R. Astron. Soc.* **63**, 117–148 (1980).
19. Rutherford, M. J. & Devine, M. J. The May 18, 1980, eruption of Mount St Helens. 3. Stability and chemistry of amphibole in the magma chamber. *J. Geophys. Res.* **93**, 11949–11959 (1988).

Acknowledgements. We thank L. Wilson for comments. E.B. was supported by a National Science Foundation graduate fellowship.

Correspondence and requests for materials should be addressed to H.M.M. (e-mail: h.m.mader@bristol.ac.uk).

The advantage of sex in evolving yeast populations

Clifford Zeyl^{*†} & Graham Bell[‡]

^{*} Center for Microbial Ecology, Michigan State University, East Lansing, Michigan 48824, USA

[‡] Department of Biology and Redpath Museum, McGill University, Montréal, Québec H3A 1B1, Canada

Sex is a general feature of the life cycle of eukaryotes. It is not universal, however, as many organisms seem to lack sex entirely¹. The widespread occurrence of sex is puzzling, both because meiotic recombination can disrupt co-adapted combinations of genes, and because it halves the potential rate of reproduction in organisms with strongly differentiated male and female gametes². Most attempts to explain the maintenance of sexuality invoke differences between parents and sexual offspring. These differences may be advantageous in novel or changing environments if new gene combinations are favoured from time to time¹. Sex would then serve to concentrate beneficial mutations that have arisen independently into the same line of descent. But in a stable environment sex might serve to concentrate deleterious mutations, so that they will be more effectively purged from the population by selection³. We have studied the effect of sex on mean fitness in experimental populations of the budding yeast *Saccharomyces cerevisiae*. Our results show that sex increases mean fitness in an environment to which the populations were well adapted, but not in an environment to which new adaptation occurred, supporting the hypothesis that the advantage of sexuality lay in the removal of deleterious mutations.

Sex is thought either to accelerate adaptation to a changing environment, or to retard the loss of adaptation in a stable environment. These are not mutually exclusive theories⁴, and it has also been suggested that sex can facilitate the establishment of beneficial mutations by disengaging them from linked deleterious mutations⁵. Although these theories have been investigated through models and simulations, few experiments have been designed to test them. The hypothesis that sex accelerates adaptation to a changing environment is supported by the results of two experiments in which recombination rates were manipulated: bacteriophage T4 populations with high recombination rates evolved proflavine resistance more rapidly than did low-recombination populations⁶, and suppression of recombination slowed the response to artificial selection on sternopleural bristle numbers in experimental populations of the fruitfly *Drosophila melanogaster*⁷.

We have studied the effect of sex in *S. cerevisiae*. Haploid yeast exist as one of two mating types α and a , which mate on contact with each other to form a/α diploids. Wild-type haploids are homothallic, but the *HO* gene required for mating-type switching in daughter cells can be inactivated, and most laboratory strains are heterothallic. When starved of nitrogen and grown on non-fermentable carbon sources such as acetate, diploids undergo meiosis and sporulation to produce haploid ascospores. The walls of ascospores are resistant to treatments such as enzymatic digestion that destroy vegetative cells. Standard laboratory techniques allow the construction of yeast genotypes that are identical except for mating type or ploidy⁸.

It has been shown that heterozygotes at the mating type (*MAT*) locus enjoy a competitive advantage over homozygotes⁹. A single episode of meiosis followed by syngamy between sibling spores (automixis or inbreeding) increased this advantage in highly

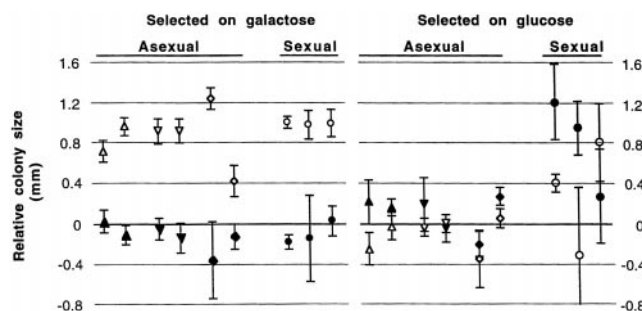


Figure 1 Growth rates of experimental yeast populations maintained on galactose or on glucose for ~400–600 generations. Mean colony sizes after three days' growth relative to a control genotype were determined for ten genotypes from each population both on the medium on which they were maintained (white symbols) and on the other medium (black symbols). Plotted values are means of the ten genotype means, with 95% confidence intervals, for replicate haploid mating type a ($\blacktriangle$), haploid mating type α ($\triangle$), diploid ($\diamond$) and sexual ($\circ$) populations.

heterozygous genotypes, but initially reduced the advantage of *MAT* heterozygosity in otherwise homozygous genotypes. But the implications of these results for an advantage of sex are not straightforward, because *MAT* heterozygosity regulates the expression of many yeast genes¹⁰ and because mitotic recombination is relatively frequent in diploid yeast¹¹.

We constructed replicate haploid mating type a , haploid mating type α , diploid populations (all asexual), and sexual populations. The constituent genotypes were constructed as isogenic strains, so that the starting populations were not only isogenic but almost homogeneous (see Methods). Replicate experimental populations were propagated by serial transfer on one of two media: 1% yeast extract, 2% peptone and 2% glucose, or an identical medium with 2% galactose substituted for glucose. Most laboratory yeast strains are maintained on glucose, and the genotypes from which our base populations were constructed divide roughly every 90 min on glucose at 25 °C. They grow only about half as fast on galactose, and induce the biochemical pathways of galactose metabolism inefficiently¹². Our starting populations were thus well adapted to glucose and relatively poorly adapted to galactose.

The sexual populations underwent eight sexual cycles, each separated by about 48 or 96 mitotic divisions for populations maintained on galactose and glucose, respectively, whereas the asexual populations experienced 8 parallel asexual cycles and were refrigerated during the 5-day period in each cycle when sexual populations were sporulating. Because mitotic reproduction occurred neither on the starvation medium used to induce meiosis nor during refrigeration, the environmental differences experienced by sexual and asexual populations have no selective effects. However, although we attempted to maintain similar population sizes in all treatments by transferring similar numbers of cells (10^5 to 10^6) at each cycle, it is possible that in some cycles the sporulation treatment resulted in smaller sexual than asexual populations.

After eight experimental cycles, vegetative fitnesses were estimated by measuring the growth rates on glucose and galactose plates of ten genotypes from each population, relative to the genotype used to construct the starting populations (which had been stored frozen in 15% glycerol).

All populations maintained on galactose adapted significantly to galactose (Fig. 1). This adaptation was specific to the galactose medium, as no such increase was seen in their growth rates when tested on glucose. We observed no differences between haploid and diploid or between sexual and asexual populations in their adaptation to galactose (Table 1).

[†] Present address: Department of Biology, Wake Forest University, P.O. Box 7325, Winston-Salem, North Carolina 27109, USA.

Table 1 ANOVA of relative colony diameters of populations maintained on galactose

Source	d.f.	Assayed on galactose MS	Expected MS	F
Sex	1	3.518	$s_e^2 + 6.8s_{g(sr)}^2 + 64.8s_{g(sa)}^2 + Q_s$	1.04
Replicate (sex)	7	3.387	$s_e^2 + 5.7s_{g(sr)}^2 + 51.3s_{g(sa)}^2$	17.19**
Genotype (replicate (sex))	79	0.197	$s_e^2 + 5.5s_{g(sr)}^2$	9.85***
Error	399	0.020	s_e^2	
Assayed on glucose				
Sex	1	0.629	$s_e^2 + 7.2s_{g(sr)}^2 + 68.8s_{g(sa)}^2 + Q_s$	1.01
Replicate (sex)	7	0.623	$s_e^2 + 7.0s_{g(sr)}^2 + 66.2s_{g(sa)}^2$	2.56*
Genotype (replicate (sex))	79	0.243	$s_e^2 + 6.8s_{g(sr)}^2$	8.68***
Error	512	0.028	s_e^2	

ANOVA of the relative colony diameters of populations maintained on galactose, with the occurrence or absence of sex as a fixed effect and replicate population and genotype as random nested effects. Among haploid populations maintained on galactose, we found no effect of mating type on growth rates on either galactose ($P > 0.3$) or glucose ($P > 0.5$). Among haploids maintained on glucose, we detected no effect of mating type on growth rates, either on glucose ($P > 0.3$) or on galactose ($P > 0.4$). Among asexual populations maintained on galactose, no effect of ploidy was detected ($P > 0.7$ on galactose, $P > 0.3$ on glucose). The same was true among asexual populations maintained on glucose ($P > 0.8$ on glucose, $P > 0.1$ on galactose). Both replicates of the haploid mating type α , haploid mating type α , and diploid treatments were therefore pooled as replicates of an asexual treatment. Values of F associated with probabilities of less than 0.05, 0.01 and 0.001 are marked with one, two or three asterisks, respectively. To compare genetic variances in sexual and asexual populations, we performed separate ANOVAs on sexual and asexual treatments maintained on each medium, with assay environment as a fixed effect and replicate population within assay environment, genotype within population, and replicate colony within genotype as nested random effects. To test for differences in genetic variance between sexual and asexual populations, we calculated F ratios using the estimates of variance attributable to genotype from each treatment. Among populations maintained on galactose, genetic variance was somewhat greater within asexual than sexual populations ($F_{36,53} = 1.83$; $0.06 > P > 0.02$), but among populations maintained on glucose, genetic variance was greater within sexual populations ($F_{30,36} = 2.453$; $P > 0.01$). MS, mean square; Q_s , effect of sex; $s_{g(sa)}^2$, effect of replicate; $s_{g(sr)}^2$, effect of genotype; s_e^2 , error.

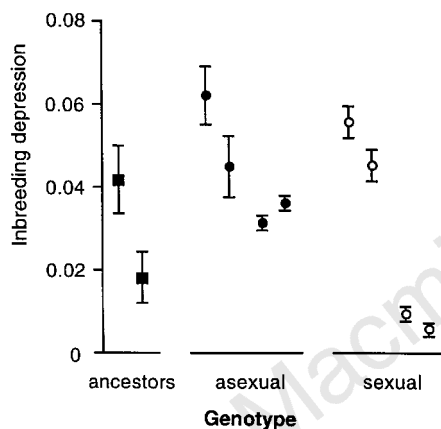


Figure 2 Inbreeding depression in ancestral genotypes and their sexual and asexual descendants. Plotted values are means and 95% confidence intervals determined as described in Methods. For both sexual and asexual genotypes, the first two values are for genotypes from replicate population 1, and the third and fourth are from replicate population 2. ANOVA with life cycle (sexual or asexual) as a fixed effect and population (replicate 1 or 2) as a random effect nested within life cycle indicated no significant effect of life cycle ($P = 0.398$), owing to the large error term (variance attributable to the effect of replicate population). Means (and 95% confidence intervals) for inbreeding depression are 0.027 (0.016–0.038) for the ancestors, 0.043 (0.034–0.052) for the asexuals and 0.029 (0.013–0.045) for the sexuals, when the results for each replicate of each genotype are pooled.

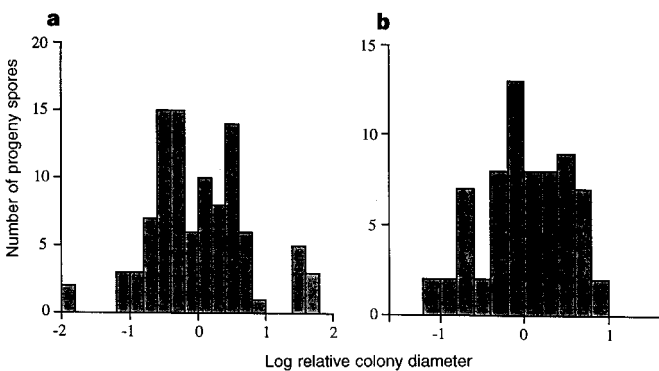


Figure 3 Frequency distributions of colony growth rates on galactose for F_1 progeny of crosses between genotypes sampled from populations that evolved on galactose, and their ancestors.

We detected no change in the performance on either medium of most of the asexual populations maintained on glucose; one diploid population grew more quickly when tested on galactose, the other diploid population grew more slowly on both media, and one of the haploid, mating type α populations grew more slowly when tested on glucose (Fig. 1). We found no effect of the test medium on the performance of any of the populations maintained on glucose, relative to the unselected control (Table 2). Thus, in contrast to the populations that evolved on galactose, those maintained on glucose acquired no detectable new adaptation to their culture medium. However, the sexual populations grew more quickly on both media than did the asexual populations (Fig. 1 and Table 2).

These two observations suggest that the superiority of the sexual populations resulted from the removal of mutations that were deleterious in both environments, rather than from the assortment of adaptive mutations into one recombinant genotype. To test this explanation, we chose two haploid genotypes from each of two sexual and two asexual haploid populations maintained on glucose, and compared the effects of inbreeding and outcrossing on the fitness of each. If the asexual genotypes carried greater numbers of deleterious mutations, and those deleterious mutations were at least partially recessive, then the fitness disadvantage of inbred versus outcrossed diploid progeny (inbreeding depression) should be greater in asexual than in sexual genotypes. We would expect any such effect to be slight, because all deleterious mutations were exposed to selection throughout the experiment in the asexual haploid population, and during the haploid phase of the sexual cycle in the sexual populations. Our results, although statistically inconclusive, are consistent with the expectation of more abundant deleterious mutations in the asexual genotypes (Fig. 2). The two ancestors of our populations carry differing and substantial loads of deleterious mutations, and the load appears to have increased in some asexual genotypes.

The maintenance of adaptation by the removal of deleterious mutations from sexual populations could occur through either of two mechanisms. First, in small asexual populations mean fitness is liable to decrease irreversibly because of the stochastic loss of the few individuals bearing the fewest mutations (Muller's ratchet^{13,14}). This does not apply to sexual populations, in which genotypes with few mutations are recreated by recombination. This might be the case, for example, in yeast strains perpetuated by the repeated transfer of single colonies. Second, the mean fitness of sexual populations may exceed that of asexual populations if each additional mutation causes a greater loss of fitness in genomes that already carry greater numbers of deleterious mutations (synergistic epistasis³). Such synergistic interactions appear to occur between ultraviolet-

Table 2 ANOVA of relative colony diameters of populations maintained on glucose

Source	d.f.	Assayed on glucose MS	Expected MS	F
Sex	1	38.919	$S_e^2 + 6.8S_{g(sr)}^2 + 62.4S_{r(rs)}^2 + Q_e$	19.34**
Replicate (sex)	7	2.012	$S_e^2 + 7.0S_{g(sr)}^2 + 65.7S_{r(rs)}^2$	3.93**
Genotype (replicate (sex))	78	0.512	$S_e^2 + 6.7S_{g(sr)}^2$	14.63***
Error	502	0.035	S_e^2	
Assayed on galactose				
Sex	1	60.978	$S_e^2 + 6.7S_{g(sr)}^2 + 59.0S_{r(rs)}^2 + Q_e$	7.30*
Replicate (sex)	7	8.356	$S_e^2 + 6.4S_{g(sr)}^2 + 57.5S_{r(rs)}^2$	20.94***
Genotype (replicate (sex))	79	0.399	$S_e^2 + 5.9S_{g(sr)}^2$	9.07***
Error	434	0.044	S_e^2	

ANOVA of the relative colony diameters of populations maintained on glucose, as in Table 1. ANOVA of the populations maintained on glucose, with assay medium glucose or galactose as a fixed effect, and replicate population within assay medium and genotype within replicate population as random effects, revealed no effect on relative colony diameter of the assay medium ($P > 0.1$).

induced and naturally accumulated deleterious mutations in *Chlamydomonas moewusii*¹⁵. We cannot distinguish between these possibilities from our results.

Although recombination would cause the same redistribution among genomes of deleterious mutations in sexual populations on galactose as in sexuals on glucose, the former showed no fitness advantage over their asexual counterparts. The increase in growth rates on galactose of all populations maintained on galactose greatly exceeded the effect of sex among populations maintained on glucose, presumably because each beneficial mutation involved in adaptation to galactose had a much greater fitness effect than each of the deleterious mutations being removed from sexual populations on glucose. This would explain why sex did not result in a similar purging of deleterious mutations from populations maintained on galactose: mutations that increased growth rates on galactose increased rapidly in frequency, overcoming the weaker selection acting upon variation in the load of deleterious mutations.

Our results show that yeast populations become adapted to a novel environment to which they were at first poorly adapted, but do not show that sex accelerates this adaptation. This would be expected if a small number of mutations, each with relatively large effect, were responsible for adaptation. When genotypes from the galactose-adapted asexual populations are crossed with their ancestors, the distribution of growth rates among the progeny is consistent with the action of single loci with major effects (Fig. 3). Although at the end of our experiment the sexual populations appeared no better adapted to galactose than the asexual populations, it is possible that this adaptation did occur earlier in the experiment in the sexual populations than in the asexuals.

So simple an adaptive process may not be typical of natural populations, especially over long periods. No experiment can address the sources of selection in the unique historical circumstances of particular organisms. Moreover, our experimental system did not include the environmental heterogeneity or antagonistic co-evolutionary interactions that are often required by theories that invoke the effect of sex in bringing together beneficial combinations of genes. Our experiment has, however, demonstrated the operation of a process that has hitherto existed only as a possibility. □

Methods

Construction and propagation of sexual and asexual populations. Haploid mating type *a*, haploid mating type *α*, diploid, and sexual base populations were composed of yeast strains yVB110, yVB114, yVB115, and equal frequencies of yVB110 and yVB114 (ref. 16), respectively. The only genetic variation in our base populations was the presence in 1% of each population of retrotransposon Ty3 and a *TRP1* or *URA3* marker, introduced into yVB114 by transformation. We detected no differences between transformant genotypes and yVB114 on either glucose or galactose ($P = 0.36$ and 0.39 , respectively, in *t*-tests of growth rates, measured as described below). All media were supplemented with 20 mg l^{-1} uracil and 40 mg l^{-1} tryptophan. Each experimental cycle began with growth for 3 days at room temperature on solid media in Petri plates, followed by transfer to 25 ml liquid medium in a 250-ml flask shaken at

250 r.p.m. Aliquots of 40 μl from each asexual population were spread on plates and refrigerated; the sexual populations were transferred to 25 ml presporulation medium (2% potassium acetate, 2% peptone, 1% yeast extract) for 24 h and then sporulation medium (1% potassium acetate, 0.1% yeast extract, 0.05% glucose) for 5 d. Unsporulated cells were killed by overnight digestion in $100 \mu\text{g ml}^{-1}$ zymolyase and 0.2% β -mercaptoethanol. Sibling ascospores were dissociated by shaking at 300 r.p.m. for 30 min in 5 ml 0.1% Tween-20 and 2 ml glass beads (0.4–0.6 mm diameter) and the ascospores were spread on plates. At this stage, the plated asexual populations were returned to room temperature and the subsequent experimental cycle began.

Assays of fitness. The relative fitness of all genotypes tested was determined either by measuring the growth rates of replicate colonies started from single cells of that genotype, or by competition experiments. The choice was determined by the number of genotypes to be scored: for assays such as the overall comparison of population mean fitnesses (Fig. 1), the number of genotypes to be tested precluded the use of the more laborious and time-consuming competition procedure. To confirm that genotypes with faster colony growth rates were also superior competitors, we did pairwise competition tests on both glucose and galactose for each of 16 pairs of genotypes taken from the final populations. In 24 of 32 cases, the genotype with the faster growth rate on a particular medium increased significantly in frequency on that medium (Fisher exact test, $P = 0.0035$).

For the overall comparison of population mean fitnesses, an aliquot from each evolved population was diluted in sterile water and thin-spread on agar. From each population, ten colonies were chosen haphazardly and streaked onto both glucose and galactose plates. After 24 h, an aliquot of each genotype was streaked onto one side of a new plate, and from each genotype 8 individual unbudded cells of typical size were picked up and placed on a $1 \text{ cm} \times 1 \text{ cm}$ grid on the plate using a Singer MSM micromanipulator. On each plate, strain yVB114, treated as above, was included as a control. After three days' growth at room temperature, the diameter of each colony was digitized using a dissecting microscope and SigmaScan software. For each colony a relative size was calculated by subtracting from its diameter the mean diameter of the yVB114 colonies on that plate. When the total variance in colony diameter is partitioned into among-genotype and within-genotype components, 89.6% of the total variance is genotypic.

The genetics of adaptation to galactose. To determine whether adaptation to the galactose medium involved few or many loci, we crossed one genotype from each population maintained on galactose with yVB110 or with yVB114, sporulated the resulting diploids, and dissected the tetrads. The ancestral strains sporulate poorly, and spore viability in these crosses was low. For each tetrad that yielded four colonies, we transferred aliquots of cells from each colony to a fresh plate and used micromanipulation to determine the mean colony size of each haploid genotype after 3 days' growth. The deviation of each genotype from the mean for that tetrad was then calculated, and the log-transformed results plotted as histograms.

Inbreeding depression. Two genotypes were chosen haphazardly from each of two sexual and two asexual populations that evolved on glucose. Each genotype tested was both outcrossed and inbred (selfed). For selfing, the genotypes were transformed with plasmid pGAL-HO, carrying a galactose-inducible copy of the *HO* gene¹⁷. Transformants were grown on galactose to induce mating-type switching and mating, and mated pairs of cells were

selected visually and isolated by micromanipulation. The diploid status of selfed genotypes was confirmed by observing that they did not mate with tester strains of either mating type. For outcrossing, each asexual genotype was mated with a genotype from a parallel sexual population (for example, genotype 1 from replicate 1 of the sexual glucose population was mated with genotype 1 from replicate 1 of either the mating-type α or mating-type α asexual glucose population, depending on the mating type of the sexual genotype). Mating cultures were spread on glucose plates, and mated pairs of cells were selected by micromanipulation. The fitnesses (w) of diploids produced by selfing and outcrossing were determined by competitions against a diploid produced by crossing yVB110 with a derivative of yVB114 carrying a URA3 genetic marker, and inbreeding depressions were calculated as $w_{\text{outcrossed}} - w_{\text{inbred}}/w_{\text{outcrossed}}$.

Received 23 April; accepted 9 May 1997.

1. Bell, G. *The Masterpiece of Nature* (University of California, San Francisco, 1982).
2. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, 1978).
3. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
4. Charlesworth, D., Morgan, T. M. & Charlesworth, B. Mutation accumulation in finite outbreeding and inbreeding populations. *Genet. Res.* **61**, 39–56 (1993).
5. Peck, J. R. A ruby in the rubbish: Beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* **137**, 597–606 (1994).
6. Malmberg, R. L. The evolution of epistasis and the advantage of recombination in populations of bacteriophage T4. *Genetics* **86**, 607–621 (1977).
7. McPhee, C. P. & Robertson, A. The effect of suppressing crossing-over on the response to selection in *Drosophila melanogaster*. *Genet. Res.* **16**, 1–16 (1970).
8. Golemis, E. A., Gyuris, J. & Brent, R. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. et al.) 13.1–13.14.17 (Wiley, Boston, 1994).
9. Birdsall, J. & Wills, C. Significant competition advantage conferred by meiosis and syngamy in the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **93**, 908–912 (1996).
10. Kunz, B. A. & Haynes, R. H. in *The Molecular Biology of the Yeast Saccharomyces* Vol. 1 (eds Strathern, J., Jones, E. W. & Broach, J.) 371–414 (Cold Spring Harbor, Plainview, 1981).
11. Mortimer, R. K. & Schild, D. in *The Molecular Biology of the Yeast Saccharomyces* Vol. 1 (eds Strathern, J., Jones, E. W. & Broach, J.) 341–370 (Cold Spring Harbor, Plainview, 1981).
12. Menees, T. M. & Sandmeyer, S. B. Transposition of the yeast retroviruslike element Ty3 is dependent on the cell cycle. *Mol. Cell Biol.* **14**, 8229–8240 (1994).
13. Muller, H. L. The relation of recombination to mutational advance. *Mutat. Res.* **1**, 2–9 (1964).
14. Crow, J. F. in *The Evolution of Sex: an Examination of Current Ideas* (eds Michod, R. E. & Levin, B. R.) 56–73 (Sinauer, Sunderland, MA, 1988).
15. De Visser, J. A. G. M., Hoekstra, R. F. & Van Den Ende, H. The effect of sex and deleterious mutations on fitness in *Chlamydomonas*. *Proc. R. Soc. Lond. B* **263**, 193–200 (1996).
16. Bilanchone, V. W. et al. Positive and negative regulatory elements control expression of the yeast retrotransposon Ty3. *Genetics* **134**, 685–670 (1993).
17. Herskowitz, I. & Jensen, R. E. Putting the *HO* gene to work: practical uses for mating-type switching. *Methods Enzymol.* **194**, 132–146 (1990).

Acknowledgements. We thank S. Sandmeyer for yeast strains and plasmids; H. Bussey for plasmids and the use of his micromanipulator; H. Reiswig for the use of digitizing equipment; D. Schoen for comments on the manuscript; and A. Kondrashov for helpful criticism. This research was supported by NSERC and FRCR grants to G.B., NSERC and FCAR scholarships to C.Z., and grants to R. Lenski.

Correspondence and requests for materials should be addressed to C. Zeyl (e-mail: zeylcw@wfu.edu).

The heritability of IQ

B. Devlin^{*†}, Michael Daniels[†] & Kathryn Roeder[†]

^{*} Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA

[†] Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

IQ heritability, the portion of a population's IQ variability attributable to the effects of genes¹, has been investigated for nearly a century, yet it remains controversial. Covariance between relatives may be due not only to genes, but also to shared environments, and most previous models have assumed different degrees of similarity induced by environments specific to twins, to non-twin siblings (henceforth siblings), and to parents and offspring. We now evaluate an alternative model that replaces these three environments by two maternal womb environments, one for twins and another for siblings, along with a common home environment. Meta-analysis of 212 previous studies shows that our 'maternal-effects' model fits the data better than the 'family-environments' model. Maternal effects, often assumed to be negligible, account for 20% of covariance between twins and 5% between siblings, and the effects of genes are correspondingly

reduced, with two measures of heritability being less than 50%. The shared maternal environment may explain the striking correlation between the IQs of twins, especially those of adult twins that were reared apart. IQ heritability increases during early childhood, but whether it stabilizes thereafter remains unclear. A recent study of octogenarians², for instance, suggests that IQ heritability either remains constant through adolescence and adulthood³, or continues to increase with age². Although the latter hypothesis has recently been endorsed⁴, it gathers only modest statistical support in our analysis when compared to the maternal-effects hypothesis. Our analysis suggests that it will be important to understand the basis for these maternal effects if ways in which IQ might be increased are to be identified.

Despite its conceptual simplicity, IQ heritability has engendered vitriolic debates throughout this century, debates that are now recurring⁵ following the publication of *The Bell Curve* by R. Herrnstein and C. Murray⁶. Part of this controversy arises because IQ heritability is not well characterized. Paradoxically, direct versus indirect analytic methods, both of which attempt to estimate genetic effects on IQ unencumbered by environmental effects, yield markedly different estimates⁷. Direct studies presumably eliminate environmental covariance by evaluating relatives raised apart; indirect studies presumably eliminate environmental covariance by contrasting results from different study designs that have complementary environmental covariance terms. Contrary to their identical objectives, the former typically produces notably larger IQ heritability estimates. We postulate that this paradox is attributable in large part to the presence of unacknowledged maternal effects on IQ.

In traditional quantitative genetic studies, the environment is divided into maternal and external constituents¹. We expand this dichotomy to take account of the realities of IQ adoption studies, lumping maternal environment and any shared external environment of adopted children into a single 'pre-separation environment'. Because the duration of shared external environment is usually small, pre-separation and maternal effects are roughly equivalent. Both are distinct from maternal inheritance, which is sometimes called a maternal effect⁸.

Our IQ analyses focus on the magnitude of the additive and non-additive genetic components¹, estimated to explain 60–85% of the variation in IQ from adult monozygotic twin studies, and the magnitude of maternal effects, usually assumed to be negligible. To estimate these effects, we analysed 212 IQ studies (or more precisely, correlations) based on 50,470 distinct pairings. The analysis included 204 correlations from studies of zero and first-degree relatives or their adoptive counterparts⁹. We supplemented this set with some new twin studies published after 1981: a study of monozygotic twins reared apart¹⁰, the Swedish adoption/twin study of aging of monozygotic twins reared together and apart and dizygotic twins reared together¹¹, and two studies of monozygotic and dizygotic adult twins reared together¹².

Each IQ correlation and related sample size is classified by kind of study (Fig. 1). We evaluate these data using a standard quantitative genetic model for the components of variance (Table 1) and Bayesian meta-analysis¹³, a standard technique for combining information across studies. Our model is built on two levels of distributional assumptions: we assume a likelihood model for the observed correlations among relatives in each type of study; and we specify a prior distribution for the parameters of the model. We assume any standardized component of variance (positive correlation) is *a priori* equally likely to lie between zero and one. These prior distributions make the Bayesian parameter estimates similar to maximum-likelihood estimates.

All correlations were Fisher-transformed and, after transformation, observed correlations from the same study design follow a normal distribution if they differ only by measurement error. However, several studies are outliers under a normal likelihood,

suggesting that the studies are more heterogeneous with respect to variability than is predicted by measurement error alone. Heterogeneity is expected because some study methods differ markedly⁹. We account for the heterogeneity by assuming that the transformed values follow a *t*-distribution with two degrees of freedom.

We fit four models to the data. The richest of these models, model IV, allows for additive and non-additive genetic effects, twin and singleton maternal effects, and external environmental effects for twins, siblings and parent–offspring. The remaining models differ from model IV in that Models I and II constrain the maternal effects to be zero, and models I and III constrain the external environmental effects to be equal.

The models are evaluated informally by comparing the average observed correlations for each of the nine study designs to the models' predicted values (Table 2). Model I predicts the correlation for the dizygotic twins raised together very poorly. Except for monozygotic twins raised apart, the predicted correlations for each study type are similar for models II–IV. The observed correlations for monozygotic twins raised apart (Fig. 1) ranges from 0.62 to 0.78, with a mean of 0.74. The most extreme predicted correlation is that obtained from model II (0.50), and the closest value is from model IV (0.74), but model III also predicts the correlation well, with a value of 0.68. Maternal effects seem to be essential for a good fit to the data.

Models III and IV estimate a large maternal effect for twins and a smaller effect for siblings (Table 3). Model III attributes 20% (95% Bayesian credible interval: 15–24%) to the former and 5% (1–8%) to the latter. The difference between these estimates is intuitively appealing. Twins share the womb concurrently, whereas siblings share the womb serially. Hence, although a mother may have similar physiological status and personal habits from one pregnancy to another, the temporal separation between progeny apparently diminishes the correlation of sibling IQ.

Table 1 Expected covariances by relationship

Relationship	Raised	Type (j)	Expected covariance
Monozygotic twins	Together	1	$\sigma_A^2 + \sigma_D^2 + \sigma_{M_T}^2 + \sigma_{ES_T}^2$
Monozygotic twins	Apart	2	$\sigma_A^2 + \sigma_D^2 + \sigma_{M_T}^2$
Dizygotic twins	Together	3	$\frac{1}{2}(1+r)\sigma_A^2 + \frac{1}{2}\sigma_D^2 + \sigma_{M_T}^2 + \sigma_{ES_T}^2$
Siblings	Together	4	$\frac{1}{2}(1+r)\sigma_A^2 + \frac{1}{2}\sigma_D^2 + \sigma_{M_S}^2 + \sigma_{ES_S}^2$
Siblings	Apart	5	$\frac{1}{2}(1+r)\sigma_A^2 + \frac{1}{2}\sigma_D^2 + \sigma_{M_S}^2$
Midparent/child	Together	6	$\sigma_A^2 + \sigma_{ES_P}^2$
Single-parent/child	Together	7	$\frac{1}{2}(1+r)\sigma_A^2 + \sigma_{ES_P}^2$
Single-parent/child	Apart	8	$\frac{1}{2}(1+r)\sigma_A^2$
Adopting parent/child	Together	9	$\sigma_{ES_P}^2$

Midparent refers to the average of the two parents' IQ. Covariances are denoted by σ^2 with subscripts for genetic and environmental effects: *A*, additive genetic; *D*, non-additive genetic; *G*, total genetic; *M_T*, maternal (twins); *M_S*, maternal (siblings); *ES_T*, *ES_S* and *ES_P* are environments for twins, siblings and parent/child, respectively; *T*, total. Standardized covariances, σ_i^2/σ_i^2 , are denoted by σ_i^2 . Assortative mating is factored into the model by the correlation of IQs between mates, *r*. Following ref. 9, we take *r* to be 0.33 (s.d. = 0.03). Reanalysing the data with *r* = 0.33 ± 2 s.d. had no notable effect on our findings.

Table 2 Posterior means for IQ correlations by study type

Relationship	Raised	Type	Model				
			0	I	II	III	IV
Monozygotic twins	Together	1	0.85	0.85	0.85	0.85	0.85
Monozygotic twins	Apart	2	0.74	0.68	0.50	0.68	0.74
Dizygotic twins	Together	3	0.59	0.46	0.59	0.59	0.60
Siblings	Together	4	0.46	0.46	0.44	0.44	0.44
Siblings	Apart	5	0.24	0.28	0.23	0.27	0.28
Midparent/child	Together	6	0.50	0.51	0.52	0.51	0.50
Single-parent/child	Together	7	0.41	0.43	0.40	0.39	0.40
Single-parent/child	Apart	8	0.24	0.25	0.23	0.22	0.21
Adopting parent/child	Together	9	0.20	0.18	0.17	0.17	0.18

Column 0 contains the weighted average of the observed correlations, and columns I–IV contain the predicted values of these correlations from models I–IV. The predicted correlations are obtained through a Bayesian simulation procedure that evaluates integrals numerically¹⁴.

Remarkably, the covariance attributed to environmental components is also roughly equal for models II–IV (Table 3). The notable difference among the models is how the covariance resulting from shared environments is partitioned.

The models are compared formally using Bayes factors^{15,16} to obtain their posterior probabilities. *A priori* each model is assumed equally likely. The best model is model III, with a posterior probability of 0.95, followed by model IV, with probability 0.05, and models I and II, with probabilities near zero. Hence model II¹⁷, which is commonly used to describe the quantitative genetics of IQ, is rejected when it is compared to model III, which involves a common home environment and a maternal environment that varies depending on whether the pair are twins or not. Model III is favoured over model II by a posterior factor of almost 10,000 to 1, primarily on account of model II's fit to the observed data from monozygotic twins reared apart: its predicted value is about six standard deviations from the mean. Model IV, which combines features of models II and III, seems to be overparameterized and its environmental components make little sense (Table 3).

Model III's parameter estimates (Table 3) define estimates of IQ heritability. However, the concept of heritability has narrow-sense and broad-sense interpretations. Narrow-sense heritability accounts for additive genetic effects only; broad-sense heritability encompasses both additive and non-additive genetic effects. This distinction is often lost in the 'IQ debates', even though it is critical to the social implications of IQ heritability¹⁸. Based on model III, broad- and narrow-sense heritability estimates are 48% and 34%, respectively; both agree closely with estimates from studies using standard quantitative genetic models^{19,20} and indirect estimation methods^{7,21}.

An IQ enigma has been developed⁷ by contrasting the typically larger direct heritability estimates with complementary, smaller indirect estimates. For example, an indirect estimate of broad-

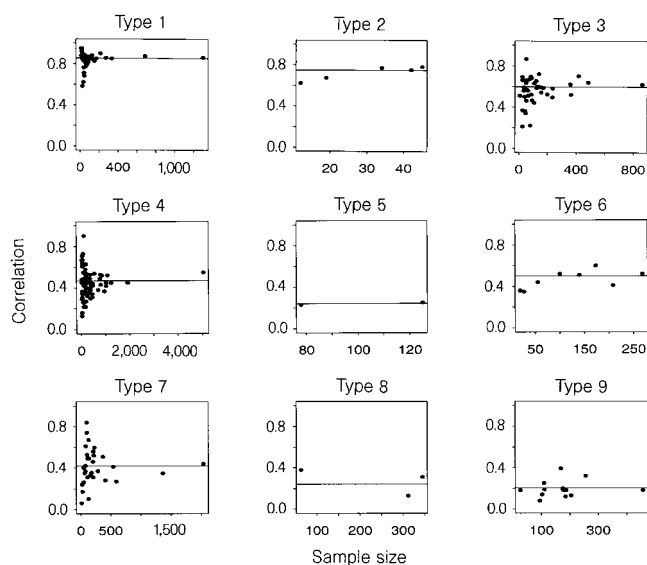


Figure 1 IQ correlations for certain study designs (see Table 1 for types) plotted versus sample size. The horizontal line indicates the mean correlation. Note the scatter about each mean is consistent with decreasing variance with increasing sample size, suggesting that much of the variance in the IQ correlations is due to sampling variance. Four type 4 studies, with missing sample size, are assigned the median sample size of 201. We analyse all study types⁶ with zero or first-degree relatives and their adoptive counterparts with the exception of adopted/adopted siblings. Adopted/adopted studies are excluded because the observed correlations are extremely variable relative to the other sets of correlations, and are inconsistent with the remaining data. If these data are included in our analysis, there is no appreciable difference in variance components estimates.

Table 3 Posterior means and 95% credible intervals for standardized covariances

Effect	Notation	Model			
		I	II	III	IV
Additive genetic	σ_A^2	0.33 (0.27, 0.39)	0.35 (0.28, 0.41)	0.34 (0.27, 0.40)	0.32 (0.26, 0.39)
Non-additive genetic	σ_D^2	0.34 (0.31, 0.38)	0.15 (0.10, 0.21)	0.15 (0.09, 0.20)	0.15 (0.10, 0.21)
total genetic	σ_G^2	0.68 (0.64, 0.71)	0.50 (0.45, 0.56)	0.48 (0.43, 0.54)	0.47 (0.42, 0.52)
Maternal (twins)	$\sigma_{M_T}^2$			0.20 (0.15, 0.24)	0.27 (0.18, 0.35)
Maternal (siblings)	$\sigma_{M_S}^2$			0.05 (0.01, 0.08)	0.06 (0.00, 0.17)
Environment (common)	σ_{ES}^2	0.17 (0.13, 0.21)		0.17 (0.13, 0.21)	
Environment (twins)	$\sigma_{ES_T}^2$		0.35 (0.30, 0.41)		0.11 (0.05, 0.19)
Environment (siblings)	$\sigma_{ES_S}^2$		0.21 (0.17, 0.25)		0.17 (0.06, 0.24)
Environment (parent/child)	$\sigma_{ES_P}^2$		0.17 (0.13, 0.21)		0.18 (0.14, 0.23)

To obtain the posterior distribution of the standardized variances, we used the Gibbs sampler with a Metropolis step¹⁴. Posterior means were obtained by numerical integration of these distributions, as were the credible intervals for the parameter estimates.

sense heritability is obtained by multiplying twice the difference between monozygotic and dizygotic twin correlations (Table 2). For these data (Table 3), the broad-sense heritability estimate of 0.52 is slightly greater than that obtained from model III, which is expected because the contrast overestimates heritability by about $0.5\sigma_D^2$ (see Table 1). The correlation for monozygotic twins reared apart yields a much larger direct estimate of 0.74. The difference between estimates can be reduced by subtracting the maternal effect from the direct estimate, yielding a value of 0.54.

Twice the observed correlation for siblings reared apart yields another direct, biased estimate of broad-sense heritability; the bias is $-1/2\sigma_D^2$, estimated from model III to be -7.5% . This direct estimate, 0.48, is much lower than the direct estimate obtained from the monozygotic twins reared apart, but it is similar to the indirect estimate and model III's estimate. The similarity to model III's estimate can be attributed to the balance between the bias and the 5% sibling maternal effect, which the direct estimates ignores. Thus maternal effects may resolve the IQ enigma by rectifying disparate direct and indirect heritability estimates.

It is not difficult to gather evidence for the importance of maternal environment. There is substantial brain growth *in utero*, and the brain has 70% of its final mass within a year of birth. Additionally, various studies indicate that IQ can be affected by prenatal environment: IQ is positively correlated with birth weight even after controlling for gestational age^{22,23}; twins, who usually weigh less than singletons, average 4–7 points lower on IQ tests²⁴; some dietary supplements can raise IQs^{25,26}; alcohol, drug and cigarette consumption may lower IQs^{27,29}; and lead exchange from mother to fetus may reduce IQ³⁰.

In contrast, the large direct estimates derived from studies of adult twins raised apart have been explained by the conjecture that, as twins get older, IQ becomes more heritable. Evidence has been provided to support this 'age hypothesis'^{4,12}, although the results are inconclusive and relationship unclear.

To compare the age- and maternal-effects hypotheses, we related the IQ correlations to subject age whenever possible. We classified the studies by age into three categories, namely youths (<14 years), adolescents (14–18 years) and adults (>18 years), based on median or mean subject age for each study; we assigned 71% of the twin and sibling studies. Because they are cross-generational, assigning parent–offspring studies was problematic: in one treatment we included all such studies as adolescents, and in another we excluded all such studies except for parent/child apart. We then extended model II. This age-effect model allowed heritability to increase with age by two adjustments: a multiplicative increase in the genetic components of variance, namely $b^{-1/2}$ for youths, b^0 for adolescents, and b^2 for adults; and a multiplicative decrease in the shared environment components of variance, namely $c^{1/2}$ for youths, c^0 for adolescents, and c^{-2} for adults. For b and $c > 1$, these multiplicative functions have forms consistent with those hypothesized in ref. 12. Our prior distribution for b and c favoured the age hypothesis slightly by putting uniform mass between $(1/\sqrt{2}, \sqrt{2})$.

The parameter estimates for b (1.13, [1.05–1.19]) and c (1.21, [1.03–1.38]), based on the analysis of all types of study, are consistent with the age hypothesis. Because Bayes factors (BF) impose a penalty for added complexity we also considered a one-parameter age-effect model with $b = c$. This model is favoured over the two-parameter model ($BF = 2.5$), and is also favoured to a small degree over model II ($BF = 11$, $b = 1.10$, [1.04–1.16]). Similar results are obtained when parent–offspring studies are excluded. However, no version of model II is competitive with model III by Bayes factor. Moreover, age-effects models fail to fit the data better than a simpler model that invokes maternal effects.

Our statistical analyses cannot, of course, be considered definitive. Age may affect IQ heritability, even though the age model does not gain much support from our analyses. Moreover, our analyses do not preclude other, unmodelled factors, such as cultural inheritance and interaction between genes and the environment, from having important effects on IQ. Despite these caveats though, several important conclusions emerge from our results.

Adoption designs are a popular means of estimating IQ heritability. Associated analyses, however, usually assume negligible maternal effects. By contrast, our results show that 20% of twin and 5% of sibling covariance may be attributable to maternal effects. These results have two implications: a new model may be required regarding the influence of genes and environment on cognitive function; and interventions aimed at improving the prenatal environment could lead to a significant increase in the population's IQ. Moreover, some of Herrnstein and Murray's conclusions⁶ regarding human evolution such as the development of cognitive castes and IQ dysgenics, arise from their belief that IQ heritability is at least 60%, and is probably closer to the 80% values obtained from adoption studies. Our results suggest far smaller heritabilities: broad-sense heritability, which measures the total effect of genes on IQ, is perhaps 48%; narrow-sense heritability, the relevant quantity for evolutionary arguments because it measures the additive effects of genes, is about 34%. Herrnstein and Murray's evolutionary conclusions are tenuous in light of these heritabilities¹⁸. □

Received 12 February; accepted 17 June 1997.

- Falconer, D. S. *Introduction to Quantitative Genetics* (Longman, New York, 1981).
- McClern, G. E. *et al.* Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science* **276**, 1560–1563 (1997).
- Gottesman, I. I. Twins: en route to QTLs for cognition. *Science* **276**, 2522–1523 (1997).
- Task Force for Scientific Affairs. *Intelligence: Knowns and Unknowns* (American Psychological Association, Washington DC, 1995).
- Devlin, B., Fienberg, S., Resnick, D. & Roeder, K. (eds) *Intelligence, Genes and Success: Scientists Respond to The Bell Curve* (Springer, New York, 1997).
- Herrnstein, R. J. & Murray, C. *The Bell Curve: Intelligence and Class Structure in American Life* (Free Press, New York, 1994).
- Plomin, R. & Loehlin, J. C. Direct and indirect IQ heritability studies: a puzzle. *Behav. Genet.* **19**, 331–342 (1989).
- Rose, R. J. *et al.* Data from kinships of monozygotic twins indicate maternal effects on intelligence. *Nature* **283**, 375–377 (1980).
- Bouchard, T. J. Jr & McGue, M. Familial studies of intelligence: a review. *Science* **250**, 223–228 (1990).
- Bouchard, T. J. Jr, Lykken, D. T., McGue, M., Segal, N. L. & Tellegen, A. Sources of human psychological differences: the Minnesota study of twins reared apart. *Science* **250**, 223–228 (1990).
- Pedersen, N. L., Plomin, R., Nesselroade, J. R. & McClearn, G. E. A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychol. Sci.* **3**, 346–353 (1992).

12. McGue, M., Bouchard, T. J. Jr, Iacono, W. G. & Lykken, D. T. in *Nature, Nurture and Psychology* (eds Plomin, R. & McClearn, G. E.) 59–76 (American Psychological Association, Washington DC, 1993).
13. National Research Council. *Combining Information. Statistical Issues and Opportunities for Research* (National Academy Press, Washington DC, 1992).
14. Smith, A. F. M. & Roberts, G. O. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Stat. Soc. B* **55**, 3–24 (1993).
15. Kass, R. E. & Raftery, A. E. Bayes Factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
16. Verdinelli, I. & Wasserman, L. Computing Bayes factors by using a generalization of the Savage-Dickey density ratio. *J. Am. Stat. Assoc.* **90**, 614–618 (1995).
17. Plomin, R., DeFries, J. C. & McClearn, G. E. *Behavioral Genetics: A Primer* (Freeman, New York, 1990).
18. Daniels, M., Devlin, B. & Roeder, K. in *Intelligence, Genes and Success: Scientists respond to The Bell Curve* (eds Devlin, B., Fienberg, S., Resnick, D. & Roeder, K.) (Springer, New York, 1997).
19. Chipuer, H. M., Rovine, M. J. & Plomin, R. LISREL modeling: genetic and environmental influences on IQ revisited. *Intelligence* **14**, 11–29 (1990).
20. Rao, D. C., Morton, N. E., Lalouel, J. M. & Lew, R. Path analysis under generalized assortative mating. II. American IQ. *Genet. Res.* **39**, 187–198 (1982).
21. McCartney, K., Harris, M. J. & Bernieri, F. Growing up and growing apart: a developmental meta-analysis of twin studies. *Psychol. Bull.* **107**, 226–237 (1990).
22. Lynn, R. & Hattori, K. The heritability of intelligence in Japan. *Behav. Genet.* **20**, 545–546 (1990).
23. Churchill, J. A. The relationship between intelligence and birth weight in twins. *Neurology* **15**, 341–347 (1965).
24. T. Husén, T. Intra-pair similarities in the school achievement of twins. *Scand. J. Psychol.* **4**, 108–144 (1963).
25. Harrell, R. F., Woodyard, E. & Gates, A. I. *The Effects of Mothers' Diets on the Intelligence of Offspring* (Teacher's College, New York, 1955).
26. Rush, D., Stein, Z., Susser, M. & Brody, N. in *Diet in Pregnancy: A Randomized Controlled Trial of Nutritional Supplements* (eds Rush, D., Stein, Z. & Susser, M.) 88–131 (Liss, New York, 1980).
27. Reinisch, J., Sanders, S. A., Mortensen, E. L. & Rubin, D. B. In utero exposure to phenobarbital and intelligence deficits in adult men. *J. Am. Med. Assoc.* **274**, 1518–1525.
28. Streissguth, A. P., Barr, H. M., Sampson, P. D., Darby, B. L. & Martin, D. C. IQ at age 4 in relation to maternal alcohol and smoking during pregnancy. *Dev. Psychol.* **25**, 3–11 (1989).
29. Olds, D. L., Henderson, C. R. Jr & Tatelbaum, R. Intellectual impairment in children of women who smoke cigarettes during pregnancy. *Pediatrics* **93**, 221–227 (1994).
30. Baghurst, P. A. et al. Environmental exposure to lead and children's intelligence at the age of seven years. *N. Engl. J. Med.* **327**, 1279–1284 (1992).

Acknowledgements. We thank T. J. Bouchard Jr for compilations of IQ data, and J. C. Loehlin, M. McGue, S. Fienberg and R. Kass for comments on a previous draft of the manuscript. This work was supported by grants from the National Science Foundation and the National Institutes of Health.

Correspondence and requests for materials should be addressed to B.D. (e-mail: devlinb@msx.upmc.edu).

Second-order fear conditioning prevented by blocking NMDA receptors in amygdala

Jonathan C. Gewirtz & Michael Davis

Departments of Psychology & Psychiatry, Yale University, Ribicoff Research Facilities of the Connecticut Mental Health Center, 34 Park Street, New Haven, Connecticut 06508, USA

Antagonists of NMDA (N-methyl-D-aspartate)-type glutamate receptors disrupt several forms of learning^{1–8}. Although this might indicate that NMDA-receptor-mediated processes are critical for synaptic plasticity, there may be other mechanisms by which NMDA-receptor antagonism could interfere with learning^{9–12}. For instance, fear conditioning would be blocked by microinfusion of the NMDA-receptor antagonist AP5 (D,L-2-amino-5-phosphonovalerate) into the basolateral amygdala^{6,13,14} if AP5 inhibited routine synaptic transmission, thereby reducing the ability of stimuli to activate amygdala neurons^{15,16}. In second-order fear conditioning^{17,18}, the reinforcer is a fear-eliciting conditioned stimulus rather than an unconditioned stimulus. Expression of conditioned fear is amygdala-dependent^{19,20} and so provides a behavioural assessment of the ability of the reinforcer to activate amygdala neurons in the presence of AP5. We report here that intra-amygdala AP5 actually enhances expression of conditioned fear to the conditioned stimulus that provides the reinforcement signal for second-order conditioning. Nevertheless, acquisition of second-order fear conditioning is completely blocked. Our findings strongly support the view that NMDA receptors are critically involved in synaptic plasticity.

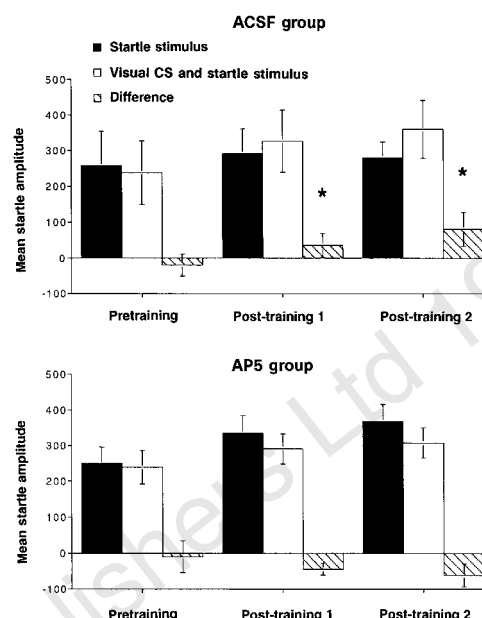


Figure 1 Fear-potentiated startle to the second-order CS in ACSF (top) and AP5 (bottom) groups tested before second-order conditioning training, and after two (post-training 1), and three (post-training 2) sessions of second-order conditioning training. Fear-potentiated startle (hatched bars, $\pm$ s.e.m.) was calculated as the difference between startle amplitude in the presence (white bars, $\pm$ s.e.m.) and absence (black bars, $\pm$ s.e.m.) of the visual CS. No infusion was given immediately before tests. An analysis of variance (ANOVA) revealed a significant group $\times$ session interaction, $F(2, 34) = 3.8$, $P < 0.05$. In the ACSF group, fear-potentiated startle increased significantly relative to pretraining baseline in post-training tests 1 and 2; values of $t(8)$ were 2.4 and 2.6 respectively, $P < 0.05$. In the AP5 group, the CS actually tended to inhibit startle relative to the pretraining baseline in post-training tests 1 and 2, although this effect was not significant; $t(9)$ was 0.7 and 1.3, respectively.

Pavlovian fear conditioning seems to involve the convergence of inputs representing the conditioned stimulus (CS) and the unconditioned stimulus (US) in the basolateral amygdala²¹. Microinfusion of NMDA receptor antagonists into the amygdala blocks the acquisition of fear conditioning, measured using the fear-potentiated startle response^{6,13}. This is consistent with data from other behavioural models suggesting that NMDA receptor activation is critical for the synaptic plasticity underlying several forms of memory^{1–8}. Alternatively, NMDA receptor blockade might interfere with the ability of the CS or the US to activate cells in the amygdala^{15,16}. It is unlikely that the ability of the CS to activate amygdala neurons is impaired, because local infusion of NMDA antagonists, in contrast to AMPA antagonists¹⁹, does not interfere with the expression of fear conditioning previously acquired^{6,13}. However, the ability of the US (typically foot shock) to activate amygdala neurons is not behaviourally testable in this learning protocol.

In pavlovian second-order fear conditioning, a second-order CS acquires conditioned fear by being paired with a first-order CS that has been paired previously with foot shock. In this case, the ability of the reinforcer to activate the amygdala can be assessed by the degree to which the first-order CS elicits conditioned fear. Thus, second-order conditioning can be used to evaluate whether an NMDA-receptor antagonist blocks acquisition of conditioned fear or interferes with synaptic transmission of the reinforcement signal in the amygdala.

To test this, male albino rats were implanted with bilateral cannulae aimed at the basolateral amygdala. In four sessions of

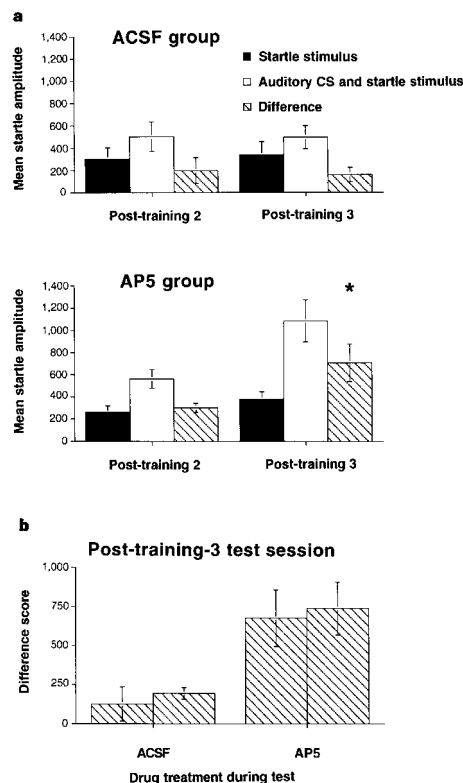


Figure 2 a, Fear-potentiated startle to the first-order auditory CS when rats were tested with no drug infused at the time of test (post-training 2), and when AP5 or ACSF vehicle were infused into the amygdala immediately before the test (post-training 3). The final test (post-training 3) was conducted after an additional session of first-order conditioning training. Fear-potentiated startle (hatched bars, $\pm$ s.e.m.) was calculated as the difference in amplitude of the startle response elicited at the start of the test session in the absence of the auditory CS (black bars, $\pm$ s.e.m.), and the startle response elicited in the presence of the auditory CS (white bars, $\pm$ s.e.m.). Intra-amygdala AP5 infusion significantly enhanced fear-potentiated startle in the final test. This was confirmed by a significant group $\times$ test session interaction $F(1, 9) = 5.2, P < 0.05$. Individual t -tests revealed that fear-potentiated startle was greater in the group infused with AP5 than in the group infused with ACSF in post-training test 3, $t(9) = 3.0, P < 0.01$, whereas the same groups did not differ significantly in the previous test, $t(9) = 1.0$. Furthermore, there was a tendency for fear-potentiated startle to be greater in the AP5 group in post-training test 3 (that is, in the presence of AP5) than in post-training test 2, $t(5) = 2.4, P = 0.06$. **b**, Fear-potentiated startle in the first and second halves of the final test session. There was no reduction of fear-potentiated startle in either group within the session, $F = 1.2$, and no group $\times$ session half interaction, $F < 1$, indicating that the enhancement of fear-potentiated startle by AP5 was not due to a blockade of extinction²³.

first-order training, an auditory CS was paired with foot shock. On the basis of a test conducted between the third and fourth training sessions, subjects were assigned to two groups with equivalent levels of fear-potentiated startle to the auditory CS. Subjects then received three sessions of second-order training, in which a visual CS was paired with the trained auditory CS. No auditory CS–shock pairings were given during these sessions. However, to reduce extinction of first-order conditioning, a fifth first-order training session was included between the second and third second-order training sessions. Immediately before each second-order training session, one group ($n = 10$) was infused with 12.5 nmol AP5 (the minimal dose effective in blocking acquisition of first-order conditioning^{6,13,14}) dissolved in artificial cerebrospinal fluid (ACSF), and a control group ($n = 9$) with the same volume of ACSF vehicle alone. Fear-potentiated startle was tested in separate sessions (without drug infusion) conducted after the second and third second-order training sessions.

The control subjects infused with ACSF alone showed acquisition of second-order conditioning, indicated by an increase in fear-potentiated startle across test sessions (Fig. 1). In contrast, subjects infused with AP5 showed no evidence of acquisition of second-order condition. In fact, the second-order visual CS produced a moderate inhibition of startle, which appeared to increase across sessions. These data indicate that AP5 blocked the acquisition of conditioned fear to the second-order visual CS.

The increase in inhibition of startle by the visual CS across test sessions seen in the AP5 group might indicate the acquisition of some form of inhibitory learning that is not blocked by intra-amygdala AP5. Alternatively, this inhibition may be related to the elevation in baseline startle across test sessions, which is somewhat greater in the absence than in the presence of the visual CS (for example, because of external inhibition¹⁷). This increase, although nonsignificant, $F(2, 18) = 2.6, P < 0.1$, may reflect an enhancement of conditioned fear to contextual cues²² learned during first-order conditioning sessions when AP5 was not given. Extinction of contextual fear in the shock-free second-order conditioning training sessions, an effect that would be blocked by intra-amygdala AP5

given on these sessions²³, could account for the fact that the same elevation in baseline startle was not seen in the ACSF group.

To assess the effect of infusion of AP5 on the expression of first-order conditioning, fear-potentiated startle to the auditory CS was tested after a further first-order conditioning session. Immediately before test, some subjects were infused with AP5 ($n = 6$) or ACSF vehicle ($n = 5$), retaining the group assignments used earlier in second-order training. As seen in Fig. 2a, both groups showed first-order conditioning, $F(1, 4) = 9.7, P < 0.05$ for the ACSF group, and $F(1, 5) = 28.7, P < 0.01$ for the AP5 group. Indeed, fear-potentiated startle was actually greater in the group infused with AP5, even though the two groups had shown equivalent levels of fear-potentiated startle in the previous test session, $F(1, 9) = 5.2, P < 0.05$. As seen in Fig. 2b, this enhancement was not attributable to a blockade of extinction of conditioned fear during the test session²³ because AP5 enhanced fear-potentiated startle equally in the first and second halves of the test session, and neither group showed extinction under these test conditions, $F = 1.2$. This enhancement of fear-potentiated startle is consistent with previous reports using both visual and auditory CSs^{6,13}. In contrast, the expression of conditioned freezing acquired to contextual cues is blocked by intra-amygdala AP5 (ref. 16). This suggests that amygdala pathways mediating explicit and contextual CS information are differentially dependent on NMDA receptor-mediated glutamatergic transmission in the amygdala.

The results of this study reduce the number of viable interpretations of the effects of NMDA receptor antagonists on fear conditioning. First, expression of fear conditioning using an auditory CS was enhanced, not impaired, at a dose of AP5 that blocked acquisition of fear conditioning⁶. This argues against the possibility that NMDA antagonists block acquisition of fear conditioning by reducing general levels of cell excitability in the amygdala¹⁶. Second, the same dose of AP5 that enhanced expression of first-order conditioning totally blocked acquisition of second-order conditioning. This indicates that the blockade of second-order conditioning was not caused by a reduction in the effectiveness of the reinforcer in activating amygdala neurons. If AP5 had no influence on synaptic

plasticity, an enhancement of first-order conditioned fear would be expected actually to support greater acquisition of second-order conditioning, because the strength of second-order conditioning depends on the strength of first-order conditioning previously established^{17,18}. Third, the enhancement of the expression of first-order conditioning by AP5 makes it unlikely that the blockade of the acquisition of second-order conditioning was a state-dependent effect. Hence, our results strongly support the view that NMDA receptors in the amygdala are critically involved in the synaptic plasticity underlying fear conditioning.

The blockade of second-order fear conditioning has further implications regarding the specific nature of the involvement of the basolateral amygdala in learning and memory. First, in contrast to foot-shock information, the pathways carrying second-order reinforcer (auditory CS) information to the basolateral amygdala^{24,25}, and their relative dependence on NMDA-receptor activation²⁶, have been well characterized. Intra-amygdala application of AP5 blocks activation of neurons in the lateral nucleus of the amygdala normally produced by stimulation of the auditory thalamus^{15,26}. In contrast, neuronal activation through stimulation of auditory cortex, a second route by which an auditory CS could access the amygdala, is not disrupted by AP5 (ref. 26). Given that expression of first-order fear conditioning, which serves as the reinforcer for second-order fear conditioning, is not disrupted by AP5, input of auditory information critical for both functions may well activate neurons of the basolateral amygdala via cortico-amygdala rather than thalamo-amygdala projections.

Our results also suggest a role of NMDA receptors in the basolateral amygdala in fear conditioning based on two different forms of learning. Whereas first-order conditioning involves the formation of an association between representations of the stimuli paired during training (that is, the CS and US, 'stimulus-stimulus' learning), the parameters used to produce second-order conditioning in this study promote the formation of associations between the second-order CS and the conditioned response (fear) elicited by the first-order CS (that is, 'stimulus-response' learning)¹⁸. The blockade of acquisition of both first-^{6,13} and second-order fear conditioning suggests that a common neural mechanism—namely, activation of NMDA receptors in the amygdala—is critically involved in the plasticity that underlies these two distinct forms of fear conditioning.

In contrast, in conditioning situations other than fear conditioning, NMDA receptors in the basolateral amygdala may be essential for acquisition based on secondary reinforcement, but not for acquisition based on primary reinforcement. Second-order appetitive conditioning is blocked by excitotoxic lesions of the amygdala, whereas, in contrast to fear conditioning²⁰, first-order appetitive conditioning is left intact²⁷. Similarly, local infusion of AP5 into the basolateral amygdala blocks acquisition of taste-potentiated odour conditioning, but not acquisition of a first-order taste aversion². Thus, the basolateral amygdala seems to be critical for learning that involves a reinforcer with conditioned motivational value, irrespective of whether that reinforcer signals impending reward or punishment^{28,29}. Outside the laboratory, stimuli with intrinsic reinforcing value, such as a noxious US, are likely to be experienced less frequently than stimuli with conditioned reinforcing value. Hence, the blockade of higher-order fear conditioning by AP5 suggests a broader involvement of amygdala NMDA receptor-mediated processes in emotional learning than would be suggested by the blockade of first-order conditioning alone. □

Methods

Surgery. Under sodium pentobarbital anaesthesia, 22 male Sprague Dawley rats were stereotactically implanted with 22-gauge cannulae aimed bilaterally at the basolateral amygdala (BLA). The cannulae were secured with dental acrylic anchored to four jeweller's screws fixed in the skull. Cannula patency was maintained using internal stylets that protruded 1 mm beyond cannula tips.

Three subjects were excluded from data analysis, either because a cannula was clogged, or because histological examination indicated that cannula placement was further than 1 mm from the BLA.

Apparatus and stimuli. For measurement of the startle reflex and delivery of all stimuli, subjects were individually housed in a stabilimeter device³⁰. The startle-eliciting stimuli were 95 dB, 50-ms bursts of white noise (rise-decay time: 5 ms), against a background white noise of 55 dB SPL (A scale). The auditory CS was a 6.7-s, 70-dB white noise that was band-pass filtered, with the low and high passes both set at 2,000 Hz. A 3.7-s visual CS was provided by an 8-W fluorescent light bulb (rise-decay time, 100 μ s; 800 foot-lamberts intensity). Scrambled, 0.5-s, 1.2-mA foot shocks were delivered through the floor bars of the stabilimeter.

Behavioural procedures. On first-order training sessions, subjects were presented with 10 initial startle stimuli, followed by 20 presentations of the auditory CS. On 16 presentations, the interval between onset of the auditory CS and foot shock was equally often 0.2, 2.2, 4.2 or 6.2 s. On the remaining 4 occasions, the CS was presented without foot shock. On second-order sessions, following the initial startle stimuli, subjects were presented with 10 pairings of auditory and visual CSs. The onset of the auditory CS occurred 3.5 s after the onset of the visual CS. In both first- and second-order sessions, the mean interval between successive presentations of the auditory CS was 150 s (range: 90–210 s).

In test sessions, following 20 initial presentations of the startle stimulus to habituate the startle reflex, startle was further elicited in the presence and absence of the CSs. Startle stimulus onset occurred at the same times after CS onset at which reinforcement had occurred in training: that is 3.5 s after visual CS onset, and 0.2, 2.2, 4.2 and 6.2 s after auditory CS onset. The interval between successive startle stimuli was 30 s. Fear-potentiated startle was calculated by subtracting mean startle amplitude in the absence of the CS from mean startle amplitude in the presence of the CS. Startle in the presence of the auditory CS was calculated as the mean amplitude of startle elicited 2.2, 4.2 and 6.2 s following CS onset.

Drug. Bilateral amygdala infusions were made with 28-gauge cannula injectors connected by polyethylene tubing to 10- μ l Hamilton syringes fitted to a Harvard apparatus infusion pump. 12.5 nmol D,L-2-amino-5-phosphonovale- rate (AP5) dissolved in 0.3 μ l ACSF was infused through each injector at a rate of 0.1 μ l min⁻¹. Injectors were left in place for 2 min following infusion.

Received 25 February; accepted 5 May 1997.

- Bannerman, D. M., Good, M. A., Butcher, S. P., Ramsey, M. & Morris, R. G. M. Distinct components of spatial learning revealed by prior training and NMDA receptor blockade. *Nature* **378**, 182–186 (1995).
- Hatfield, T. & Gallagher, M. Taste-potentiated odor conditioning: Impairment produced by infusion of an N-methyl-D-aspartate antagonist into basolateral amygdala. *Behav. Neurosci.* **109**, 663–668 (1995).
- Kim, J. J., DeCola, J. P., Landeira-Fernandez, J. & Fanselow, M. S. N-methyl-D-aspartate receptor antagonist APV blocks acquisition but not expression of fear conditioning. *Behav. Neurosci.* **105**, 125–133 (1991).
- Kim, M. & McGaugh, J. L. Effects of intra-amygdala injections of NMDA receptor antagonists on acquisition and retention of inhibitory avoidance. *Brain Res.* **585**, 35–48 (1992).
- Izquierdo, I. & Medina, J. H. Correlation between the pharmacology of long-term potentiation and the pharmacology of memory. *Neurobiol. Learn. Mem.* **63**, 19–32 (1995).
- Miserendino, M. J. D., Sananes, C. B., Melia, K. R. & Davis, M. Blocking of acquisition but not expression of conditioned fear-potentiated startle by NMDA antagonists in the amygdala. *Nature* **345**, 716–718 (1990).
- Burns, L. H., Everitt, B. J. & Robbins, T. W. Intra-amygdala infusion of the N-methyl-D-aspartate receptor antagonist AP5 impairs acquisition but not performance of discriminated approach to an appetitive CS. *Behav. Neural Biol.* **61**, 242–250 (1994).
- Staubli, U., Thibault, O., DiLorenzo, M. & Lynch, G. Antagonism of NMDA receptors impairs acquisition but not retention of olfactory memory. *Behav. Neurosci.* **103**, 54–60 (1989).
- Servatius, R. J. & Shors, T. J. Early acquisition, but not retention, of the classically conditioned eyeblink response is N-methyl-D-aspartate (NMDA) receptor dependent. *Behav. Neurosci.* **110**, 1040–1048 (1996).
- Saucier, D. & Cain, D. P. Spatial learning without NMDA receptor-dependent long-term potentiation. *Nature* **378**, 186–189 (1995).
- Keith, J. R. & Rudy, J. W. Why NMDA-receptor-dependent long-term potentiation may not be a mechanism of learning and memory: Reappraisal of the NMDA-receptor blockade strategy. *Psychobiology* **18**, 251–257 (1990).
- Caramanos, Z. & Shapiro, M. L. Spatial memory and N-methyl-D-aspartate receptor antagonists APV and MK-801: Memory impairments depend on familiarity with the environment, drug dose, and training duration. *Behav. Neurosci.* **108**, 30–43 (1994).
- Campeau, S., Miserendino, M. J. D. & Davis, M. Intra-amygdala infusion of N-methyl-D-aspartate receptor antagonist AP5 blocks acquisition but not expression of fear-potentiated startle to an auditory conditioned stimulus. *Behav. Neurosci.* **106**, 569–574 (1992).
- Fanselow, M. S. & Kim, J. J. Acquisition of contextual Pavlovian fear conditioning is blocked by application of an NMDA receptor antagonist D,L-2-amino-5-phosphonovale-ric acid to the basolateral amygdala. *Behav. Neurosci.* **108**, 210–212 (1994).
- Li, X. F., Phillips, R. & LeDoux, J. E. NMDA and non-NMDA receptors contribute to synaptic

- transmission between the medial geniculate body and the lateral nucleus of the amygdala. *Exp. Brain Res.* **105**, 87–100 (1995).
16. Maren, S., Aharonov, G., Stote, D. L. & Fanselow, M. S. *N*-methyl-D-aspartate receptors in the basolateral amygdala are required for both acquisition and expression of conditioned fear in rats. *Behav. Neurosci.* **110**, 1365–1374 (1996).
 17. Pavlov, I. P. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (Dover, New York, 1927).
 18. Rescorla, R. A. *Pavlovian Second-order Conditioning: Studies in Associative Learning* (Erlbaum, Hillsdale, NJ, 1980).
 19. Kim, M., Campeau, S., Falls, W. A. & Davis, M. Infusion of the non-NMDA receptor antagonist CNQX into the amygdala blocks expression of fear-potentiated startle. *Behav. Neural Biol.* **59**, 5–8 (1993).
 20. Campeau, S. & Davis, M. Involvement of the central nucleus and basolateral complex of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *J. Neurosci.* **15**, 2301–2311 (1995).
 21. Romanski, L. M., Clugnet, M.-C., Bordi, F. & LeDoux, J. E. Somatosensory and auditory convergence in the lateral nucleus of the amygdala. *Behav. Neurosci.* **107**, 444–450 (1993).
 22. Campeau, S. *et al.* Induction of the *c-fos* proto-oncogene in the rat amygdala during unconditioned and conditioned fear. *Brain Res.* **565**, 349–352 (1991).
 23. Falls, W. F., Miserendino, M. J. D. & Davis, M. Extinction of fear-potentiated startle: Blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12**, 854–863 (1992).
 24. Campeau, S. & Davis, M. Involvement of subcortical and cortical afferents to the lateral nucleus of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *J. Neurosci.* **15**, 2312–2327 (1995).
 25. Romanski, L. M. & LeDoux, J. E. Equipotentiality of thalamo-amygdala and thalamo-cortico-amygdala circuits in auditory fear conditioning. *J. Neurosci.* **12**, 4501–4509 (1992).
 26. Li, X. F., Stutzmann, G. E. & LeDoux, J. E. Convergent but temporally separated inputs to lateral amygdala neurons from the auditory thalamus and auditory cortex use different postsynaptic receptors: *In vivo* intracellular and extracellular recordings in fear conditioning pathways. *Learn. Mem.* **3**, 229–242 (1996).
 27. Hatfield, T., Han, J.-S., Conley, M., Gallagher, M. & Holland, P. C. Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J. Neurosci.* **16**, 5256–5265 (1996).
 28. Everitt, B. J. & Robbins, T. W. *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (ed. Aggleton, J. P.) 401–429 (Wiley-Liss, New York, 1992).
 29. Willick, M. L. & Kokkinides, L. Cocaine enhances the expression of fear-potentiated startle: Evaluation of state-dependent extinction and the shock-sensitization of acoustic startle. *Behav. Neurosci.* **109**, 929–938 (1995).
 30. Cassella, J. V. & Davis, M. The design and calibration of a startle measurement system. *Physiol. Behav.* **36**, 377–383 (1986).

Acknowledgements. We thank Bill Falls, whose pilot studies suggested a role for NMDA receptors in acquisition of second-order conditioning and stimulated this investigation. This research was supported by grants from AFPSR and NIMH (to M.D.) and by an NIH individual NRSA from NIMH (to J.C.G.)

Correspondence and requests for materials should be addressed to J.C.G. (e-mail: jonathan.gewirtz@yale.edu).

Transport, docking and exocytosis of single secretory granules in live chromaffin cells

J. A. Steyer, H. Horstmann & W. Almers

Max-Planck-Institut für medizinische Forschung, 69120 Heidelberg, Jahnstrasse 29, Germany

Neurons maintain a limited pool of synaptic vesicles which are docked at active zones and are awaiting exocytosis^{1–4}. By contrast, endocrine cells releasing large, dense-core secretory granules have no active zones, and there is disagreement about the size⁵ and even the existence⁶ of the docked pool. It is not known how, and how rapidly, secretory vesicles are replaced at exocytic sites in either neurons or endocrine cells. By using electron microscopy, we have now been able to identify a pool of docked granules in chromaffin cells that is selectively depleted when cells secrete. With evanescent-wave fluorescence microscopy⁷, we observed single granules undergoing exocytosis and leaving behind patches of bare plasmalemma. Fresh granules travelled to the plasmalemma at a top speed of 114 nm s⁻¹, taking an average of 6 min to arrive. On arrival, their motility diminished 4-fold, probably as a result of docking. Some granules detached and returned to the cytosol. We conclude that a large pool of docked granules turns over slowly, that granules move actively to their docking sites, that docking is reversible, and that the 'rapidly releasable pool' measured electrophysiologically represents a small subset of docked granules.

Even in chromaffin cells, a population of docked dense-core granules can be morphologically identified if chemical fixation is avoided (Fig. 1a). We measured the closest distance between granules and the plasmalemma and plotted the results as a histogram (Fig. 1b). Granules populated the cytosol at a constant density, although those touching the plasmalemma were more numerous. In terms of granules per μm^2 (Fig. 1 legend), the cytosolic density of granules (dashed line) is only 5% of the density of those touching the plasmalemma. The remaining 95% must be bound to the plasmalemma, or docked. Their number in a resting cell was $1.68 \pm 0.11 \mu\text{m}^{-2}$ (5 experiments) or $1,010 \pm 80$ in a typical cell with a $600 \mu\text{m}^2$ surface area. When cells were stimulated by elevating the external potassium ion concentration ($[\text{K}^+]$), granules were lost by exocytosis (Fig. 1c). Granules touching the plasmalemma were lost selectively (Fig. 1d); $53 \pm 11\%$ of them were lost during 2 minutes of stimulation (5 experiments). Selective depletion of docked granules over 2 min is not expected if cytosolic and docked granules reach equilibrium over that time. For instance, the 'rapidly releasable pool' of secretory vesicles in chromaffin cells⁸ and neurons^{9,10} is replenished with time constants of only 10 s. In chromaffin cells, this evidently can happen without docking new granules.

The dynamics of these morphological changes was studied in living cells²⁴ by fluorescence microscopy. Cells were allowed to attach to glass coverslips, then stained briefly with acridine orange and washed thoroughly. For viewing single granules beneath the plasmalemma, confocal microscopy had insufficient depth resolution and, most importantly, could image single granules only at illumination levels that caused rapid bleaching and phototoxicity. To image a thin layer of cytosol with less intense illumination, we used a method not previously applied in neurobiology. Fluorescence was excited selectively in a 300-nm-thin layer of cytosol immediately adjacent to the coverslip, using the evanescent wave set up by a laser beam as it suffered total reflection at the glass/cell interface⁷. Typical images (Fig. 1e, f) show scattered fluorescent spots. Because granules are by far the most abundant organelle near the plasmalemma (Fig. 1a), most spots must represent granules close to the plasmalemma where it adheres to the coverslip.

Stimulation with high $[\text{K}^+]$ caused fluorescent spots to disappear (Fig. 1f) as exocytosing granules lost their dye. Granules were lost in patches, consistent with earlier reports of localized secretion¹¹. The patches did not arise because the plasmalemma buckled and moved out of the evanescent wave, because in simultaneous interference reflection images (Fig. 1g, h) the 'footprint' of the cell on the coverglass remained dark; the footprint would have brightened had the plasmalemma detached. During a 2-min exposure to high $[\text{K}^+]$, $34 \pm 6\%$ of the granules were lost (15 cells). This is less than the 53% lost in Fig. 1b, c, presumably because the evanescent wave also excites fluorescence in some granules that do not touch the plasmalemma and hence cannot perform exocytosis. Comparison of the two values suggests that in a resting cell, $34/53 = 64\%$ of the fluorescent spots are secretory granules touching the plasmalemma.

The cells we studied had 1.3 ± 0.4 fluorescent spots per μm^2 (24 cells). If only 64% represent granules touching the plasmalemma, we see fewer such granules (0.8 per μm^2) than in electron micrographs. Although plasmalemma adhering to glass may carry fewer granules, we believe that two other factors are also important: first, we tended to select cells with lower granule density (see Methods); second, granules in electron micrographs varied 2–3-fold in diameter and hence more than 8-fold in volume. If granules carry dye and fluoresce in proportion to their volume, small ones will be overlooked if they are next to larger ones.

Four findings indicate that most of the fluorescent granules in Fig. 1e, f were lost through exocytosis. First, the effect was not seen when Mg^{2+} replaced external Ca^{2+} and 0.1 mM Cd^{2+} was added ($99 \pm 4\%$ remaining after 2 min at high $[\text{K}^+]$; 5 cells). Second, fluorescent spots disappeared abruptly from one frame to the next

(Fig. 2a, images 2, 3) even in recordings made at 1 frame per second (not shown). If instead, granules left the plasmalemma and moved into the cytosol, they would need a median time of 14 s to escape the evanescent wave and become invisible (see below). Third, a diffuse cloud of secreted dye often appeared briefly at the point where a granule disappeared (Fig. 2a, images 5, 6). Finally, recordings of single catecholamine quanta (Fig. 2c) revealed that significant secretion happened only during stimulation (Fig. 2b), as did the loss of granules (Fig. 2d). Secretion of quanta and loss of granules proceeded roughly proportionately to each other.

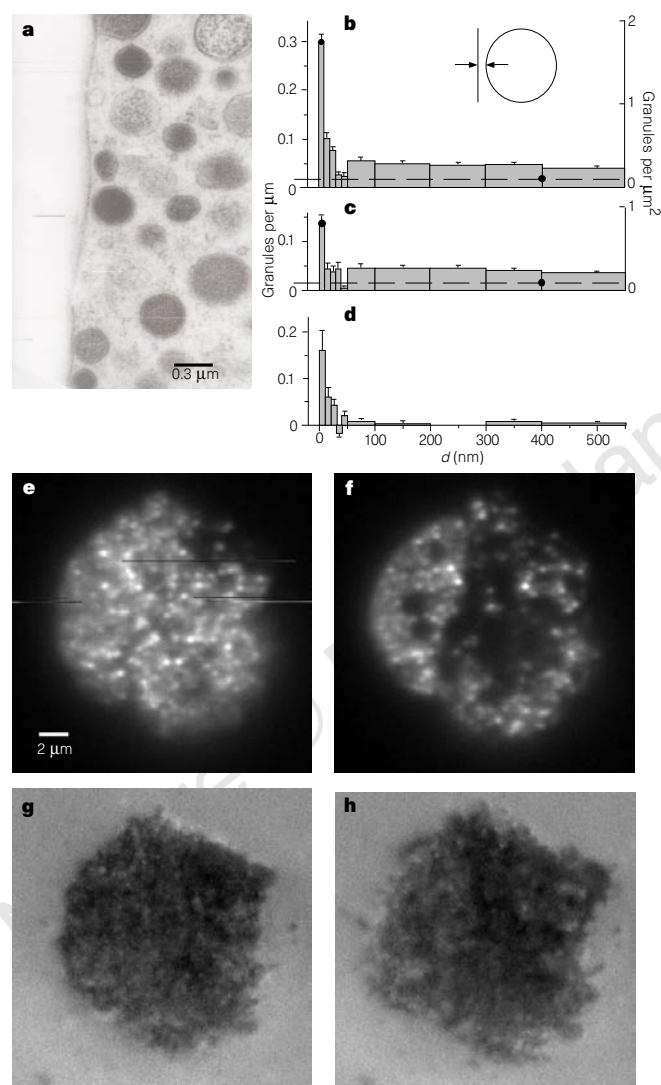


Figure 1 **a**, Unfixed, quickly frozen chromaffin cell showing plasmalemma (left) and the granules beneath. **b**, Number of granules as a function of distance d from the plasmalemma, measured as shown in the inset and calculated for 10-nm bin width. Bars give granules per μm of plasmalemma; filled circles were converted into granules per μm^2 . The filled circle at 400 nm (dashed line) is the mean per 10-nm bin for $150 < d < 600$ nm, and proportional to the cytosolic granule density ($0.11 \mu\text{m}^{-2}$ for a 10 nm layer, or $10.5 \mu\text{m}^{-3}$). **c**, As in **b**, but after 2-min stimulation in a solution with high $[\text{K}^+]$, expected to raise cytosolic $[\text{Ca}^{2+}]$ by influx through plasmalemmal calcium channels. **d**, Granules lost during the 2-min stimulation (difference between **b** and **c**). **e–h**, Light micrographs showing the region of a cell where it adheres to a glass coverslip. **e**, **f**, Evanescent-wave fluorescence micrographs: **e**, in normal buffer, and **f**, after 2 min in high $[\text{K}^+]$, as in **c**. **g**, **h** Interference reflection images taken within 1 s of **e**, **f**, respectively.

To determine how rapidly the depleted plasmalemma is repopulated with granules, we watched fluorescent spots appear after stimulation stopped (Fig. 3a, b). As dye had been absent from the bathing medium for about 1 hour, the new spots must represent stained granules that had moved from the cytosol into the evanescent wave and onto the plasmalemma (Fig. 3c). The plasmalemmal granule population was partially replenished with a time constant of 6 min. This is >10 -fold slower than at synapses, where endocytic vesicles become release-ready synaptic vesicles in only 30–60 s (hippocampal synapses¹²) or 15–30 s (frog neuromuscular junction¹³).

The appearance of a single granule beneath the plasmalemma is shown in Fig. 4a; Fig. 4b plots the time course of fluorescence at the granule site. When the granule appeared, fluorescence rose above background and increased to a plateau as the granule reached the plasmalemma. As the fluorescence grew, the granule brightened continually, never reversing its course towards the plasmalemma over 14 s. Such directed movement is not expected for a diffusion process and indicates that the granule moved actively. The granule ultimately docked successfully, because a second stimulus triggered its exocytosis (Fig. 4b, arrowhead); this happened 174 s after its arrival and 96 s after the second stimulus. Once at the plasmalemma, a granule evidently reaches fusion competence within minutes.

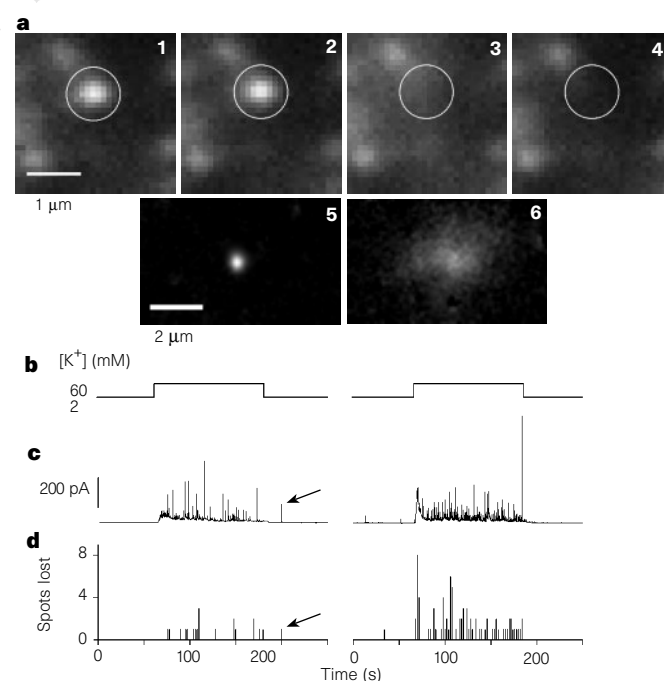


Figure 2 **a**, 1–4, Successive frames at 0.5 s^{-1} show a granule (1, 2) vanishing abruptly owing to exocytosis (2, 3); the granule and its last location are indicated by circles. Images 5, 6 are the same as 2, 3 after subtracting image 4. The summed glow of the cloud in image 6 was more than that from the granule before exocytosis. This is consistent with secretion against the coverslip, where the dye is most strongly excited by the evanescent wave. **b–d**, Behaviour of two other cells during a similar experiment. **b**, External $[\text{K}^+]$ was raised and lowered by a rapid perfusion system. **c**, Current spikes recorded by a carbon fibre near the cell; each represents a single quantum of released catecholamine²³. Secretion starts when $[\text{K}^+]$ is raised and stops when it is lowered again. **d**, Data from simultaneous video clips; number of granules lost in successive 2-s intervals plotted against time. One granule was located at the margin of the cell facing the carbon fibre, and its disappearance coincided with the release of a catecholamine quantum (arrow). Generally, however, carbon fibre and microscope sample different regions of the cell, hence the detected events lack one-to-one correspondence.

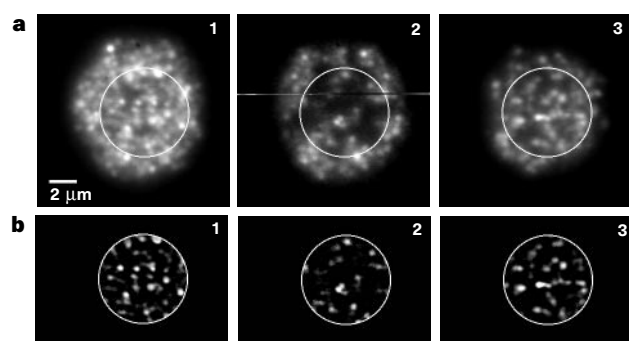


Figure 3 a, Images before (1), immediately after (2) and 20 min after the end of a 2-min stimulation with elevated $[K^+]$ (3). **b**, Area marked by the circle in **a**. To enhance contrast, we subtracted from each image a version that had been low-pass filtered at a spatial frequency of $1/\mu\text{m}$. **c**, Top, drawing showing coverslip with adherent portion of a cell; granules fluoresce (black) only within the evanescent wave (shaded area). Bottom, granule density within the circles in **b** normalized to the prestimulus value ($1.4 \pm 0.1 \mu\text{m}^{-2}$) and plotted as a function of time. High $[K^+]$ applied within rectangle. Both the decline and recovery of granule density (to $69 \pm 9\%$ in this dataset of 6 cells) varied strongly from cell to cell. To reduce effects of this variability on our estimate of the recovery time course, post-stimulus data for each cell were scaled to force the average of the last four measurements to 0.69. Continuous line, exponential giving best least-squares fit (time constant 6 min).

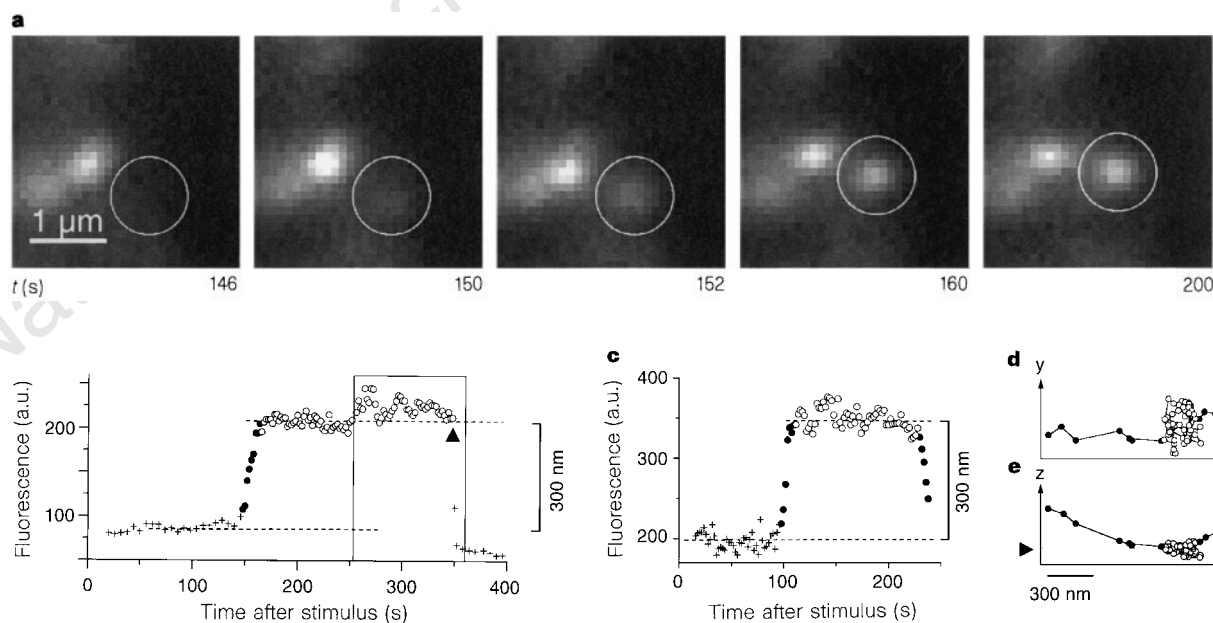
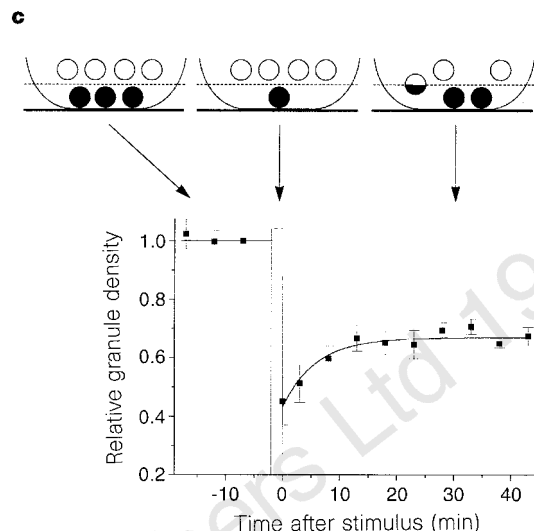


Figure 4 a, A granule appearing beneath the plasmalemma. Times shown are relative to the end of the first stimulus; circles outlining the granule are centred on its positions. **b**, Fluorescence intensity in the centre of the circles plotted against time. a.u., Arbitrary units. The initial fluorescence is background from neighbouring granules or from out-of-focus granules excited weakly by deeply penetrating photons of the evanescent wave. For the duration shown by the rectangle, $[K^+]$ was elevated for a second time. Accompanying intensity fluctuations are due to a second granule approaching and retreating laterally (images not shown). Dashed lines: fluorescence before the granule appeared (lower) and immediately before it underwent exocytosis (arrowhead); they provide an approximate calibration for

the granule distance for the plasmalemma. Filled circles, during the granule approach: this is, from the time it first appeared to when it exceeded 80% of its average plateau intensity. Open circles, during the time the granule remained bright. Crosses indicate when the granule was invisible and were measured at the locations where the granule appeared and disappeared. The first cross after the granule's disappearance is higher owing to the cloud of secreted dye. **c**, Similar analysis on another granule; symbols as in **b**. **d**, Lateral, and **e**, vertical trajectories of the granule in **c**. Positions were determined at 2-s intervals; arrowhead marks the putative position of the plasmalemma; z is the vertical coordinate calculated from distance calibration in **c**.

With plasmalemma and glass in contact (images as in Fig. 1g, h; results not shown), the approaching granule must have traversed the entire 300-nm layer where the evanescent wave makes it visible in our microscope (see Methods). With this distance calibration (dashed lines in Fig. 4b, c), the granule's vertical velocity (the component perpendicular to the plasmalemma) was 24 nm s^{-1} , averaged from first appearance to 80% peak intensity ($34 \pm 17 \text{ nm s}^{-1}$; 8 granules in 5 cells). In 14 approaching granules (6 cells), the median time to reach 80% of maximum brightness was 14 s.

Figure 4c show the result of a similar analysis in the absence of a stimulus. This granule did not exocytose, but its fluorescence remained high and nearly constant as if it, too, had reached the plasmalemma. After 100 s, the granule dimmed gradually as it moved back into the cytosol. Figure 4d plots its lateral trajectory (that is, in the plane of the plasmalemma). The granule approached in a highly directed fashion over nearly $0.5 \mu\text{m}$ (filled circles), apparently moving by means other than diffusion. If a diffusion coefficient is calculated from the average distance moved in the 2-s interval between frames in Fig. 4d, the chance of encountering the approach trajectory (filled circles) by diffusion is 4×10^{-5} . On arrival at the plasmalemma, directed motion stopped (open circles), suggesting that docking had occurred, and restarted when the granule detached again (filled circles). In all of 6 similar observations, including those shown in Fig. 4b, the lateral motility diminished on arrival at the plasmalemma and increased again if the granule detached, changing by 0.65 ± 0.15 log units or 4.5-fold (5 cells). The loss of motility at the plasmalemma indicates hindrance by an added drag force, probably due to capture by a plasmalemmal docking site. The residual apparently random lateral motion (open circles in Fig. 4d) was analysed by measuring the mean distance moved in successive 8-s intervals. It could be accounted for by a diffusion coefficient of $3.0 \pm 0.4 \times 10^{-4} \mu\text{m}^2 \text{ s}^{-1}$ ($n = 6$). The corresponding value for plasmalemma-resident granules in resting cells was similar ($3.0 \pm 0.8 \times 10^{-4} \mu\text{m}^2 \text{ s}^{-1}$; 5 cells with 2–6 granules each). Both values are 5,000 times smaller than they would be in water.

For a granule reaching the plasmalemma, its distance from the plasmalemma during approach or departure may be calculated from its intensity (Fig. 4c), and combined with the data in Fig. 4d to plot the trajectory in a vertical plane (Fig. 4e). The granule landed on the plasmalemma at an angle (Fig. 4e, filled circles), nearly stopped moving for 100 s after touch-down (open circles) and then detached and ascended back into the cytosol. Evidently, docking is reversible. Most approaching granules moved laterally as well as vertically, as in Fig. 4c–e. The average approach speed in all three dimensions was $40 \pm 16 \text{ nm s}^{-1}$, with top speeds of $114 \pm 11 \text{ nm s}^{-1}$ (9 granules in 6 cells; 22–25°C). At higher temperatures, this top speed might approach that of actin-based organelle transport in neuronal growth cones (480 nm s^{-1} ; 37°C; ref. 14). Like many other cells, chromaffin cells have a dense subplasmalemmal network of actin filaments^{15,16}.

We have tracked single granules *en route* from the cytosol to their docking sites and hence to their exocytosis. In ultrastructural studies, a typical resting chromaffin cell had 1,000 secretory granules stuck to its plasmalemma. Most could exocytose during 2 min stimulation with elevated $[\text{K}^+]$. Stimulated exocytosis was observed by evanescent-wave microscopy and seen to cause patches of empty plasmalemma. These were later repopulated with fresh granules, taking an average 6 min (22°C) to arrive. The pool of morphologically docked granules in chromaffin cells was much larger and was replenished far more slowly than the 'rapidly releasable' pool defined electrophysiologically⁸. Probably granules do not attain exocytic competence immediately after docking, and the docking complex must undergo subsequent maturation.

We conclude that granules reach the plasmalemma in a directed manner, moving actively. After arrival, granules lose lateral

motility, consistent with the ultrastructural finding that they bind to the plasmalemma. Future experiments on permeabilized cells must identify the motors that carry granules to their docking sites, and address the molecular and functional mechanism of docking. □

Methods

Bovine chromaffin cells¹⁷ were used 2 to 3 days after plating onto glass coverslips in buffer containing (in mM) 135 NaCl, 2 KCl, 5 CaCl₂, 2 MgCl₂, 10 Na-HEPES, pH 7.2, which kept the plasmalemma at a potential where calcium channels are shut and secretion does not occur. To stimulate secretion, 60 mM KCl replaced NaCl for 2 min. All experiments were done at 22–25°C; values are $\pm$ s.e.

Electron microscopy. Cells were quickly frozen after 2 min in buffer with low or high $[\text{K}^+]$, by plunging them into liquid ethane, and processed as described^{17,18}. Analysis was at 100,000-fold magnification by counting peripheral granules per length of plasmalemma, as determined with a map measuring wheel. To convert the results into granules per μm^2 of plasmalemma, they were divided by the sum of the section thickness (100 nm) plus the diameter of the structure being sampled¹⁹. This was $0.077 \mu\text{m}$ for the contact zone between the membrane of a docked granule and the plasmalemma (defined as the distance over which both remained within 10 nm; median, 194 granules), and $0.324 \mu\text{m}$ for granules located more than one granule radius from the plasmalemma¹⁷.

Light microscopy. Cells were stained for 15 min in buffer with $3 \mu\text{M}$ acridine orange at 22°C, then washed thoroughly in dye-free buffer and viewed 30–90 min later (Zeiss 100 $\times$, 1.4 NA oil objective, 16-bit back-illuminated CCD camera; Princeton Instruments). Interference reflection images (IRM's) were taken with 600 nm light coupled into the epillumination port with an appropriate short-pass dichroic mirror. To reduce stray reflection, we subtracted an image taken at the same focal plane but showing no cell. IRM images are darker the closer the plasmalemma approaches the coverslip; with our illumination aperture of about 1.2, all areas darker than background are expected to lie within 80 nm of the coverslip²⁰.

Fluorescence images were taken with evanescent-wave excitation. We selected cells whose clear and dark 'footprint' in IRM indicated close adherence to the cover glass. We rejected about half of them because fluorescent spots were not clearly distinct and blended into each other. This selection probably excluded cells with high granule densities. 488-nm light from an argon laser was passed into the microscope's epifluorescence port. An annular mask, placed in a plane conjugate to the objective's back focal plane, blocked all but the most marginal rays⁷. These leave the objective at a supercritical angle and hence suffer total reflection at the coverslip/cell interface. Such epi-illumination arrangement allowed free access to the cell, so that secretion could be measured with a carbon fibre (8 μm tip diameter at 0.7 V; ref. 21) placed 1–2 μm from the cell surface. To probe the evanescent wave, we held fluorescent 280 nm diameter beads (similar in size to chromaffin granules) at varying distances from the coverslip. This was done by placing a planoconvex lens curved-side down onto a coverslip, and letting beads adsorb to the surfaces of both. The vertical distance between beads at the two surfaces was measured with a calibrated piezoelectric focusing drive. When the objective was focused on beads at the coverslip, peak fluorescence from beads at the lens decreased roughly linearly with their distance from the coverslip. The regression line fell to zero at 311 nm (not shown). For counting granules and for tracking their paths in the plane of the membrane, images were processed as for Fig. 3b, and granule locations determined with an algorithm²² (Metamorph, Universal Imaging) that identified the centre of mass of a granule's fluorescence with a standard deviation of $<20 \text{ nm}$ (tests with fluorescent beads). To compare lateral motilities in Fig. 4c, we divided the distance covered during the granule's approach (represented by the distance between the first and last filled circle) by that covered during an interval of the same duration but starting 20 s after the point represented by the first open circle (putative arrival at the plasmalemma). When granules approached slowly, we used only the last 10 locations during the approach. The fluorescence intensity of an individual granule was measured as the average over a $0.3 \times 0.3 \mu\text{m}$ square placed in its centre. Granules were rejected if, during their stay near the plasmalemma (open circles), the root-mean square intensity fluctuation was $>10\%$ of the fluorescence increase brought by the granule's arrival.

Received 18 April; accepted 28 May 1997.

1. Brodie, K. *et al.* Syntaxin and synaptobrevin function downstream of vesicle docking in *Drosophila*. *Neuron* **15**, 663–673 (1995).
2. Pieribone, V. A. *et al.* Distinct pools of synaptic vesicles in transmitter release. *Nature* **375**, 493–497 (1995).
3. Ceccarelli, B. & Hurlbut, W. P. Vesicle hypothesis of the release of quanta of acetylcholine. *Physiol. Rev.* **60**, 396–441 (1980).
4. Hunt, J. M. *et al.* A post-docking role for synaptobrevin in synaptic vesicle fusion. *Neuron* **12**, 1269–1279 (1994).
5. Gillis, K. D. & Chow, R. H. Kinetics of exocytosis in adrenal chromaffin cells. *Sem. Cell Dev. Biol.* (in the press).
6. Trifaro, J. M. & Vitale, M. L. Cytoskeleton dynamics during neurotransmitter release. *Trends Neurosci.* **16**, 466–472 (1993).
7. Stout, A. L. & Axelrod, D. Evanescent field excitation of fluorescence by epiillumination microscopy. *Appl. Optics* **28**, 5237–5242 (1989).
8. Moser, T. & Neher, E. Rapid exocytosis in single chromaffin cells recorded from mouse adrenal slices. *J. Neurosci.* **17**, 2314–2323 (1997).
9. Stevens, C. F. & Tsujimoto, T. Estimates for the pool size of releasable quanta at a single central synapse and for the time required to refill the pool. *Proc. Natl Acad. Sci. USA* **92**, 846–849 (1995).
10. von Gersdorff, H. & Matthews, G. Depletion and replenishment of vesicle pools at a ribbon-type synaptic terminal. *J. Neurosci.* **17**, 1919–1927 (1997).
11. Schroeder, T. J., Jankowski, J. A., Senyshyn, J., Holz, R. W. & Wightman, R. M. Zones of exocytotic release on bovine adrenal medullary cells in culture. *J. Biol. Chem.* **269**, 17215–17220 (1994).
12. Ryan, T. A. & Smith, S. J. Vesicle pool mobilization during action potential firing at hippocampal synapses. *Neuron* **14**, 983–989 (1995).
13. Betz, W. J. & Wu, L.-G. Kinetics of synaptic-vesicle recycling. *Curr. Biol.* **5**, 1098–1101 (1995).
14. Evans, L. L. & Bridgman, P. C. Particles move along actin filament bundles in nerve growth cones. *Proc. Natl Acad. Sci. USA* **92**, 10954–10958 (1995).
15. Lee, R. W. & Trifaro, J. M. Characterization of anti-actin antibodies and their use in immunocytochemical studies on the localization of actin in adrenal chromaffin cells in culture. *Neurosci.* **6**, 2087–2108 (1981).
16. Cheek, T. R. & Burgoyne, R. D. Nicotine-evoked disassembly of cortical actin filaments in adrenal chromaffin cells. *FEBS Lett.* **207**, 110–114 (1986).
17. Parsons, T. D., Coorsen, J. R., Horstmann, H. & Almers, W. Docked granules, the exocytic burst, and the need for ATP hydrolysis in endocrine cells. *Neuron* **15**, 1085–1096 (1995).
18. Tse, F. W., Tse, A., Hille, B., Horstmann, H. & Almers, W. Local Ca^{2+} release from internal stores controls exocytosis in pituitary gonadotrophs. *Neuron* **18**, 121–132 (1997).
19. Weibel, E. R., Gonzague, S. K. & Scherle, W. F. Practical stereological methods for morphometric cytology. *J. Cell Biol.* **30**, 23–38 (1966).
20. Gingell, T. & Todd, I. Interference reflection microscopy. A quantitative theory for image interpretation and its application to cell-substratum separation measurement. *Biophys. J.* **26**, 507–526 (1979).
21. Chow, R. H. & von Rüden, L. in *Single-Channel Recording* (eds Sakmann, B. & Neher, E.) 245–275 (1995).
22. Ghosh, R. N. & Webb, W. W. Automated detection and tracking of individual and clustered cell surface low density lipoprotein receptor molecules. *Biophys. J.* **66**, 1301–1318 (1994).
23. Wightman, R. M. *et al.* Temporally resolved catecholamine spikes correspond to single vesicle release from individual chromaffin cells. *Proc. Natl Acad. Sci. USA* **88**, 10754–10758 (1991).
24. Terakawa, S., Fan, J. H., Kumakura, K. & Ohara-Imaizumi, M. Quantitative analysis of exocytosis directly visualized in living chromaffin cells. *Neurosci. Lett.* **123**, 82–86 (1991).

Acknowledgements. We thank J. Howard, M. Lindau, W. M. Roberts and T. Soldati for their helpful suggestions on the manuscript.

Correspondence and requests for materials should be addressed to W.A. (e-mail: almers@mfz.mpi-helidelberg.mpg.de).

Optical detection of a quantal presynaptic membrane turnover

Timothy A. Ryan*, Harald Reuter† & Stephen J Smith

Department of Molecular and Cellular Physiology, Stanford University Medical School, Stanford, California 94305, USA

† Department of Pharmacology, University of Bern, 3010 Bern, Switzerland

Exploration of the mechanisms and plasticity of synaptic transmission has been hindered by the lack of a method to measure single vesicle turnover directly in individual presynaptic boutons at isolated nerve terminals. Although postsynaptic electrical recordings have provided a wealth of invaluable basic information about quantal presynaptic processes¹, this approach has often proved difficult to apply at most central nervous system synapses^{2–6}. Here we describe the direct optical detection of single quantal events in individual presynaptic boutons of cultured hippocampal neurons. Using the fluorescent dye FM 1-43 as a tracer for presynaptic endocytosis^{7–10}, we have characterized

both evoked and spontaneous components of presynaptic function at the level of individual quanta. Our results are consistent with quantal interpretations of previous electrophysiological analyses^{1–6} and provide new information about the unitary membrane recycling event and its coupling to individual action potential stimuli, about spontaneous vesicle turnover at individual boutons, and about the numbers of vesicles recycling at individual boutons.

The strategy we used for optical detection of single presynaptic quanta builds upon earlier studies using the fluorescent dye FM 1-43 as a marker for synaptic vesicle turnover^{7–10}. So far there has been no direct evidence demonstrating that uptake and release of such a vesicle marker can occur in quantized units corresponding to single vesicles. Our approach has been to include refinements designed for the measurement of very weak presynaptic fluorescence signals. Single-molecule fluorescence detection^{11–14} has shown that detection of weak signals is limited primarily by the presence of background fluorescence in instruments and specimens. We used confocal laser scanning microscopy to minimize instrumental autofluorescence background, and sparse monolayer cell cultures and the briefest possible staining exposures to FM 1-43 to minimize specimen fluorescence backgrounds.

A scheme of our experimental protocol is shown at the top of Fig. 1. Figure 1a is a Nomarski image illustrating the site of a putative axo-dendritic synaptic junction (arrow). Figure 1b, c shows fluorescence images of the same field, acquired sequentially after FM 1-43 loading with a single action potential (AP) stimulus (image A in the schematic): both panels show a small area of bright fluorescence localized to the junction site—measurement reproducibility is excellent and is generally sustained for hours at unstimulated boutons. Figure 1d, e shows reduction in fluorescence produced by a long unloading train of APs (image B). The fluorescence signal ΔF is defined as the difference in bouton fluorescence measurements before and after unloading (that is, images A–B). This ΔF signal presumably represents the amount of dye taken up into a releasable vesicular pool during the specified load. A second series (Fig. 1f–h) with more stimulation during loading (20 AP) verifies the presence of a viable presynaptic bouton at the site of interest. The 20 AP load resulted in 16-fold greater uptake of releasable dye (Fig. 1i).

To test for the postulated quantization of FM 1-43 signals, we measured fluorescence values across a large number of minimally stimulated individual boutons. Figure 2a represents individual measurements from a representative experiment in which just 1 AP was stimulated: each vertical set of points represents three separate measurements from the same bouton (fluorescence after loading but before unloading, black circles; the residual non-releasable fluorescence measured after extensive unloading, white circles; and the intrinsic fluorescence of the synapse measured before loading (F_i), squares); solid, vertical lines represent ΔF , and dotted lines represent amplitudes of the residual fluorescence signal F_R . These data show that releasable fluorescence (ΔF) is ~10 times greater than F_i , but that F_R magnitude is in the same range as the minimal ΔF signals. Because F_R became visible after even the briefest dye exposures, regardless of whether any APs are fired, this weak, non-releasable fluorescence probably reflects a residual membrane staining with no special relationship to synaptic vesicle recycling. This interpretation of F_R is further supported by the lack of any significant bouton-by-bouton correlation between F_R and ΔF values (data not shown).

Figure 2b shows a histogram of the same ΔF values as in Fig. 2a. This histogram shows a distinct peak at a ΔF amplitude of 63 photoelectrons. There are also hints of secondary and tertiary peaks at double and triple that amplitude, but clearly the number of observations is inadequate for much certainty here. By pooling observations from several experiments, however, a multimodal response amplitude distribution becomes much more obvious. Figure 3a shows a frequency histogram of the different values of

* Present address: Department of Biochemistry, Cornell University Medical College, 1300 York Avenue, New York, New York 10021, USA.

ΔF measured from 409 individual boutons stimulated by 1, 2 or 3 APs. (The 2- and 3-AP stimulus runs were included to favour the possible occurrence of more multiple-quantum events.) The histogram shows a prominent peak centred at 60 photoelectrons, with higher-order peaks appearing precisely at integral multiples. This clearly multimodal response distribution is exactly as would be predicted from the hypothesis that releasable staining is taken up in quantal units corresponding to individual synaptic vesicles.

The solid line in Fig. 3a is a curve generated as the sum of four gaussian functions centred about exact integer multiples of the mean of the first peak (see Methods). The variance within individual peaks appears to be accounted for entirely by the expected photon shot-noise statistics associated with each measurement (see Methods). The peak spacing, F_Q , used to fit this histogram provides a factor that can now be used to calibrate any ΔF measurement in terms of an actual number of presynaptic quanta recycled ($N_Q = \Delta F/F_Q$). This calibration also indicates that the releasable fluorescence illustrated in Fig. 2b–e corresponds to a single quantum.

Our results may underestimate the numbers of quanta turned over by minimal AP stimuli. Dye exposures were limited here to ~ 7 s to minimize fluorescence backgrounds, whereas earlier work has shown that substantial AP-evoked endocytosis may trail stimulation with a half-time of as long as 20 s. Thus, this low-noise procedure may fail to label some number of late-endocytosing vesicles¹⁵.

It has long been known¹ that presynaptic quantal release can occur in a spontaneous as well as an evoked fashion. To characterize spontaneous vesicle turnover under our experimental conditions, we carried out experiments in which no APs were stimulated and in which calcium influx during any possible spontaneous APs (already minimized, if not eliminated by extracellular CNQX) were blocked by adding 100 μ M extracellular Cd^{2+} . A dye exposure of 20 s gave the results shown in Fig. 3b, in which the frequency histogram of ΔF was measured at 212 different boutons from 5 different experiments. The distribution of ΔF values shows a prominent peak at 63 photoelectrons, and a second peak at twice that value. The positions of these peaks are similar to those obtained with AP stimulation (Fig. 3a). These data thus indicate that the endocytic uptake associated with spontaneous vesicular release and recycling is identical to that evoked by AP firing. These measurements also provide a means to assess frequencies of spontaneous release at individual boutons. Spontaneous recycling frequencies (f_s) were determined by measuring N_Q for a fixed dye exposure period, τ , from a population of identified presynaptic boutons. The frequency f_s , calculated for each bouton as N_Q/τ , varied considerably across boutons as well as different experiments. Average values of f_s ($\langle f_s \rangle$) determined from individual experiments spanned a tenfold range (0.01 s^{-1} – 0.1 s^{-1}).

This low frequency of spontaneous release indicates that spontaneous vesicle recycling must make only a very minor contribution

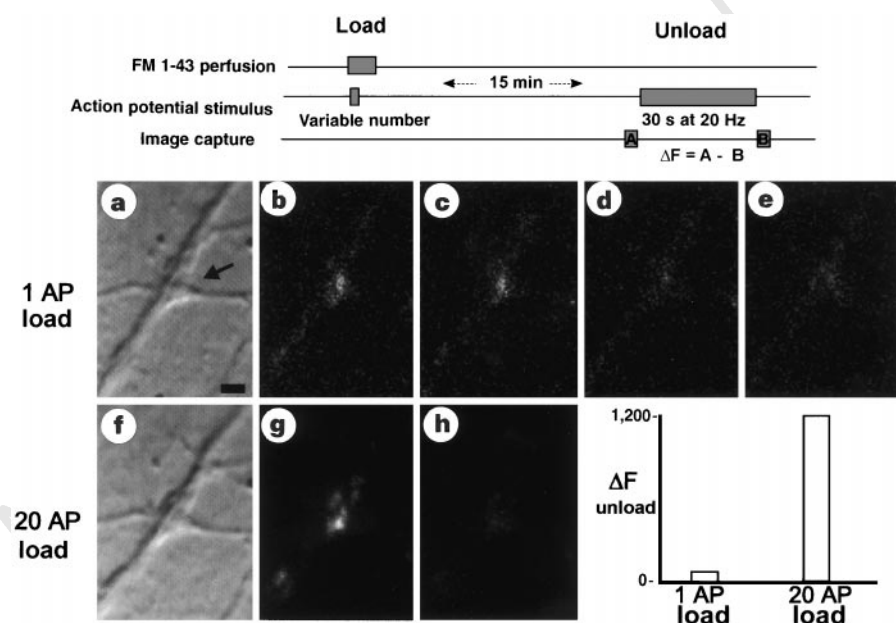


Figure 1 FM 1-43 visualization of vesicle turnover response to a single AP. Top, recycling vesicles were labelled by exposure to 15 μ M extracellular FM 1-43 (Load) during AP stimulus. Thirty seconds of 20-Hz stimulation in a dye-free solution then released vesicular dye (Unload). **a**, A Nomarski image of hippocampal culture showing putative axo-dendritic synapse (arrow). Scale bar, 2 μ m. **b**, Fluorescence of the same field following a single AP dye load; **c**, as **b**, but repeated 30 s later; **d**, **e**, as **b**, **c**, but after dye unloading; **f**, as **a**, at the beginning of a second loading/unloading sequence; **g**, as **b**, following a 20 AP load. The fluorescence signal is much brighter (although displayed here with lower-contrast gain) and additional fluorescent puncta are now apparent. **h**, As **g**, after unloading; **i**, comparison of ΔF (calibrated in photoelectrons) associated with 1 and 20 AP stimuli.

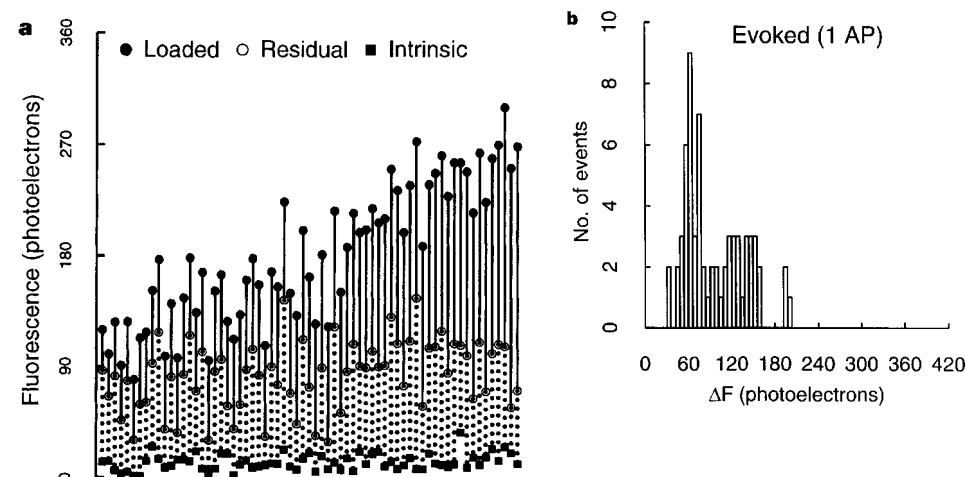


Figure 2 Fluorescence with a 1 AP loading stimulus, as measured from 67 individual boutons in a single experiment. **a**, Three measurements were made from each bouton: the intrinsic fluorescence before any dye exposure, F_I (filled squares); fluorescence after loading, F_L (filled circles); fluorescence after unloading, F_R (open circles). Length of solid vertical line corresponds to releasable fluorescence ΔF ; length of the dotted line to the dye-related residual fluorescence. **b**, A frequency histogram of the same ΔF data as in **a**.

to our measurement of evoked quantal activity. Figure 3c confirms this point: the number of quanta loaded at individual boutons by sequential stimulus episodes consisting of one and then five APs is in approximate proportion to the number of APs fired. If spontaneous quantal activity was important in dye uptake under these conditions, a disproportionately large signal would be expected from the single AP stimulus.

As a further test of the key expectation that F_Q should reflect the nature of the quantal event independently of its probability of occurrence, we reduced the probability of evoked release by using a low-calcium (0.5 mM) high-magnesium (3.5 mM) superfusate¹. Figure 3d shows that the resulting frequency histogram is still multimodal and that the quantal spacing F_Q is essentially unchanged from that observed at normal calcium/magnesium concentrations (2 mM each), even though the probability of release per AP was considerably reduced.

Estimates of the number of quanta composing the total recycling vesicle pool, N_{total} , were obtained by measuring the ratio $\Delta F_S/F_Q$ at individual boutons, where ΔF_S is the signal obtained for saturating load conditions (400 AP load). Figure 4a shows a representative frequency distribution of different N_{total} measured at 215 boutons

from three experiments and indicates that $\langle N_{\text{total}} \rangle = 127$. The fact that this number agrees closely with ultrastructural estimates of hippocampal vesicle pool sizes^{16,17} provides additional support for our conclusion that presynaptic ΔF quanta correspond to single presynaptic vesicles.

Measurements of quantal turnover per AP for brief-stimulus trains (7 AP), as well as N_{total} at the same synapses, were also made. For these experiments, longer dye exposures were used (Fig. 4) to minimize the possible underestimation of quantal turnover number, at the cost of modest but tolerable increases in background fluorescence noise. Data from such experiments indicate that the average fractional turnover of the entire pool per AP obtained in a single experiment was $0.47 \pm 0.03\%$ ($n = 54$), in close agreement with previous estimates obtained using macroscopic release assays of FM-1-43-labelled vesicles at hippocampal boutons¹⁸. A comparison of the quantal turnover to N_{total} at each synapse in a given experiment reveals a strong correlation between these two quantities (Fig. 4b). These results suggest that a single AP results in the uptake of between 0 and 3 quanta per bouton, with the average turnover across all competent boutons being 0.80 ± 0.08 quanta per AP. (Because of the possibility that turnover is still underestimated

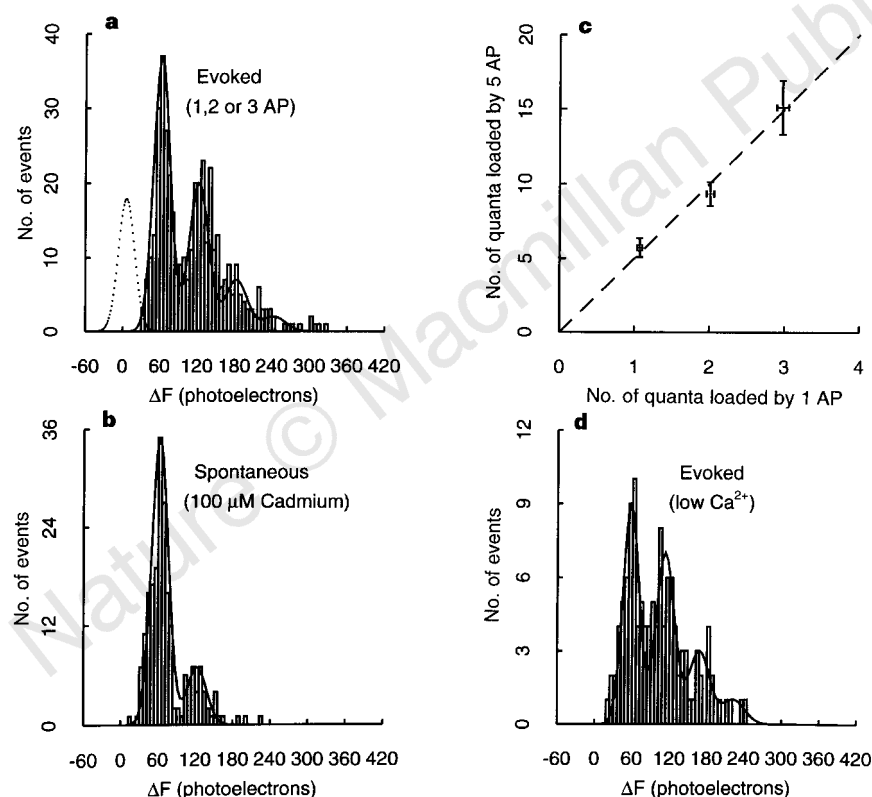


Figure 3 Minimally stimulated fluorescence signal amplitudes. **a**, Histogram of ΔF values from 8 experiments (1, 2 or 3 AP load at 0.3 Hz). Solid line is a multiple gaussian with peaks at equal 60-photoelectron intervals and widths based on a photodetection noise model (see Methods). Dotted line is a single gaussian fitted to the background noise distribution (see Methods). **b**, Histogram of ΔF values from spontaneous vesicle turnover (212 boutons, 5 experiments). Solid-line fit from equation (1) (α_3, α_4 , both 0; $\mu = 63$ photoelectrons). **c**, Scaling of quantal turnover with AP number. Sequential measurements of ΔF from 1 AP and 5 AP loads of the same bouton were binned by nearest integral value of N_Q for 1 AP load ($n_1 = 44$, $n_2 = 28$, $n_3 = 7$, errors, s.e.m.). Dotted line has a slope of 5. **d**, Histogram for uptake evoked by 4, 5, 6 or 7 AP loads in low- Ca^{2+} , high- Mg^{2+} saline (119 synapses, 4 experiments). Solid line from equation (1), with $\mu = 58$ photoelectrons.

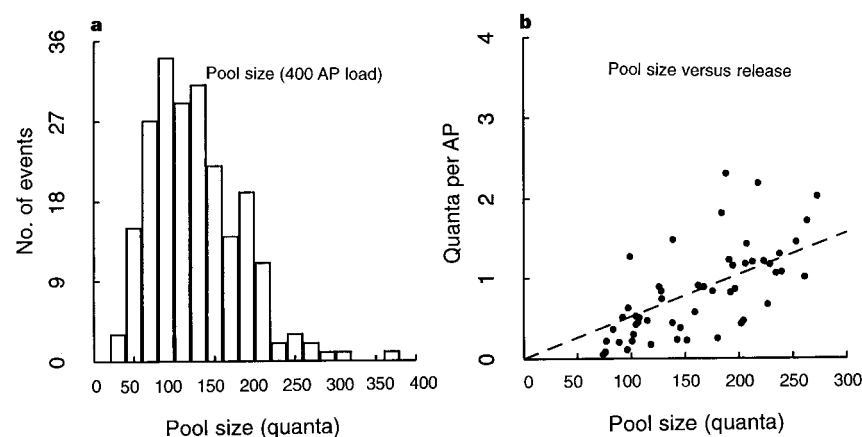


Figure 4 Estimates of total recycling pool size and fractional release per AP. **a**, Pool size distribution. The distribution of pool sizes (N_{total}) is shown in 3 experiments measuring each of 215 boutons. **b**, Quantal turnover/AP versus pool size. Three sequential measurements of ΔF were used to determine F_Q (1 AP), N_Q /AP (7 AP) and N_{total} (400 AP) in a single experiment. For the latter two experiments, dye exposures were extended to 15 and 60 s respectively. The scatter plot of N_Q /AP versus the total pool size N_{total} indicates a correlation between these two quantities (correlation coefficient, 0.68).

owing to the brevity of dye exposure relative to endocytosis delays, these number should be considered as lower limits.) If a strict, one-to-one coupling of unitary endocytic and exocytic events is assumed, our results indicate that a single AP may sometimes lead to multiple exocytic events at a single bouton^{19–21}. It is possible, however, that boutons with higher release rates actually comprise multiple active zones, and that evoked release per active zone might be limited to a maximum of one vesicle.

In summary, we have shown that dye uptake occurs in quantal steps during evoked and spontaneous presynaptic function, that such steps have the magnitude expected for uptake into a single synaptic vesicle, and that such endocytosis can proceed in minimum steps of one single quantum. Our findings strongly support endocytic vesicle recycling models of presynaptic function in central synapses. Until now, the idea that presynaptic endocytosis occurs in steps corresponding to single vesicles has been based on ultrastructural observations. The observed temporal proximity of single-quantum dye uptake events to single-AP stimuli also suggests that unitary exocytic and endocytic events are coupled closely in some way. Our optical method for presynaptic quantal analysis has several features that are likely to make it a very important complement to postsynaptic electrophysiological approaches: first, it is inherently insensitive to postsynaptic factors such as receptor saturation and modulation that can complicate interpretation of postsynaptic responses^{5,6,20,22}; second, it has an inherent ability to discriminate function at individual boutons, which is difficult with electrophysiological measurements^{19,23,24}; third, it offers high throughput because tens or hundreds of individual boutons can be analysed in parallel in a single imaging experiment; and finally, it is both convenient and amenable to long-term experimentation because no recording pipette needs to be introduced. We expect that these features will prove highly useful in the ongoing analysis of presynaptic function and plasticity.

Note added in proof: In a recent report²⁵ another example of multimodal FM1-43 distribution is indicated in the inset of Fig. 2. □

Methods

Most experimental details have been described¹⁸. Briefly, hippocampal CA1–CA3 regions were dissected from 4-day-old Sprague–Dawley rats, dissociated and plated onto Matrigel-coated coverslips and cultured for 3–5 weeks before use. Coverslips were mounted in a laminar-flow microscope chamber and continuously superfused with a saline solution consisting of 119 mM NaCl, 2.5 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 25 mM HEPES (buffered to pH 7.4), 30 mM glucose. Synaptic transmission was blocked by 10 μM 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; Research Biochemicals) 50 μM D,L-2-amino-5-phosphonvaleric acid (AP-5; Research Biochemicals) to prevent recurrent firing of unstimulated APs. Low-Ca²⁺ saline consisted of 0.5 mM CaCl₂ and 3.5 mM MgCl₂, but otherwise identical components. The dye-loading solution contained 15 μM FM 1-43 (Molecular Probes). For experiments to resolve individual quanta, dye exposures were kept brief (typically 7 s) to minimize background membrane staining. FM1-43 does not affect frequencies or amplitudes of spontaneous excitatory post-synaptic currents (EPSCs) measured in separate electrophysiological experiments (A. Bouron and H.R., unpublished results), indicating that the dye does not alter presynaptic function. APs were stimulated by passing 1-ms current pulses between electrodes placed at opposite ends of the perfusion chamber. In control experiments in which cells were stained with Fluo-3 AM, calcium imaging showed that neurons in our hippocampal cultures fired APs in response to 1-ms current pulses at a threshold field of ~4 V cm⁻¹ in both normal and low-Ca solutions. For all FM 1-43 experiments, field strengths of 10 V cm⁻¹ were used to ensure reliable action-potential firing. Simultaneous fluorescence and Nomarski imaging was performed with a modified Bio-Rand MRC 500 laser scanning unit coupled to a Zeiss IM-35 inverted microscope and a Nikon 40× 1.3 NA Fluor objective. Regions of culture coverslips were selected for imaging using a requirement that most boutons be in focus within a single image plane, as determined using both fluorescence and Nomarski images. If a few boutons lay outside a single focal plane, those boutons were eliminated from subsequent

analyses. For each image measurement, 2 frames were averaged using a dwell time of ~1.5 μs per pixel, and a sampling frequency of ~150 nm per pixel. The attenuated beam from an argon ion laser (200 microwatts at 488 nm) was used to slightly underfill the objective pupil. Fluorescence signals were corrected for dark current and calibrated in photoelectrons, using a factor obtained from fluctuation analysis of the photomultiplier output at different illumination intensities. Comparison of images like those in Fig. 1b, c with similarly acquired images of a subresolution fluorescent bead (50 nm diameter; data not shown) demonstrated that the apparent spatial extent of the synaptic fluorescence punctum is exactly as expected for a subresolution object such as a 50-nm vesicle, given the resolution limits of this confocal microscope.

Image analysis. Fluorescence measurements were made by averaging signals over a 4 × 4 pixel (0.36 μm²) region centred over bouton images. Boutons were selected for analysis of minimal-stimulus response based on their identification using long loading and unloading trains carried out at the end of each experiment. Boutons for which the fluorescence image shifted laterally by more than 4 pixels during the unloading (~10–20% of boutons in the field, probably resulting from motion of residual fluorescence) were also rejected. Boutons whose ΔF signal was within 1.5 standard deviations of the mean of the background noise distribution (≤26 photoelectrons) were considered to be below threshold and were eliminated, as were any boutons whose initial fluorescence values saturated the 8-bit digitization scale. The number of boutons scored as below-threshold decreased in proportion to the magnitude of the stimulus in a given experiment. Photodetection noise and the difference signal attributable to photobleaching were estimated from analysis of image-to-image variations across serially repeated images of minimally stained boutons. Noise data from 97 boutons obtained in this way were fitted with a single gaussian with s.d. = 14.5 photoelectrons. The photobleaching rate was ~5% per image. ΔF histograms were fitted with sums of gaussian curves with equally spaced peaks and widths set to approximate the variance expected from a poisson model of photodetection noise statistics as follows:

$$\sum_{i=1}^4 \frac{\alpha_i}{\sqrt{2\pi(i+2)\mu}} e^{-\frac{(x-\mu)^2}{2(i+2)\mu}} \quad (1)$$

where μ is the integral spacing of the means, α_i is the amplitude of each gaussian, and x is the position along the ΔF axis. The gaussian approximation of the underlying poisson process is valid as μ > 10. The standard deviation of each gaussian, i, was set to $\sqrt{(i+2)\mu}$, the expected dispersion arising from the two sequential measurements: image A consists of fluorescence from i stained vesicles, each contributing μ photoelectrons, plus a residual equal to ~μ photoelectrons; image B consists only of the residual of μ photoelectrons. As the histograms consist of a difference measurements of A–B, the total variance is the sum of variances from each measurement ((i × μ + μ) + μ). The unconstrained variables in this case are the individual amplitudes for each peak (α_i), and the centre of the first peak, μ. In contrast to classical quantal analysis¹, in which measurements on a given synapse are repeated serially, the peak amplitudes cannot be derived from poisson statistics as we are performing only a single measurement on each of an array of synapses of unknown statistical homogeneity.

Received 26 February; accepted 13 May 1997.

- Katz, B. *The Release of Neural Transmitter Substances* (Liverpool University Press, 1969).
- Redman, S. Quantal analysis of synaptic potentials in neurons of the central nervous system. *Physiol. Rev.* **70**, 165–198 (1990).
- Korn, H. & Faber, D. S. Quantal analysis and synaptic efficacy in the CNS. *Trends Neurosci.* **14**, 439–445 (1991).
- Kullmann, D. M. Quantal analysis using maximum entropy noise deconvolution. *J. Neurosci. Meth.* **44**, 47–57 (1992).
- Edwards, F. A. Anatomy and electrophysiology of fast central synapses lead to a structural model for long-term potentiation. *Physiol. Rev.* **75**, 759–787 (1995).
- Stevens, C. F. Quantal release of neurotransmitter and long-term potentiation. *Neuron* **10** (suppl.), 55–63 (1993).
- Betz, W. J. & Bewick, G. S. Optical analysis of synaptic vesicle recycling at the frog neuromuscular junction. *Science* **255**, 200–203 (1992).
- Betz, W. J., Mao, F. & Bewick, G. S. Activity-dependent staining and destaining of living motor nerve terminals. *J. Neurosci.* **12**, 363–375 (1992).
- Ryan, T. A. et al. The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron* **11**, 713–724 (1993).
- Reuter, H. Measurements of exocytosis from presynaptic nerve terminals reveal heterogeneous inhibition by Ca²⁺ channel blockers. *Neuron* **14**, 773–779 (1995).
- Betz, E. & Chichester, R. J. Single molecules observed by near-field scanning optical microscopy. *Science* **262**, 1422–1425 (1993).
- Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).

13. Mertz, J., Xu, C. & Webb, W. W. Single-molecule detection by two-photon-excited fluorescence. *Optics Lett.* **20**, 2532–2534 (1995).
14. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93**, 6264–6268 (1996).
15. Ryan, T. A., Smith, S. J. & Reuter, H. The timing of synaptic vesicle endocytosis. *Proc. Natl Acad. Sci. USA* **93**, 5567–5571 (1996).
16. Harris, K. M. & Stevens, J. K. Dendritic spines of CA 1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.* **9**, 2982–2997 (1989).
17. Malgaroli, A. LTP expression: hanging like a yo-yo? *Semin. Cell Biol.* **5**, 231–241 (1994).
18. Ryan, T. A. & Smith, S. J. Vesicle pool mobilization during AP firing at hippocampal synapses. *Neuron* **14**, 983–989 (1995).
19. Stevens, C. F. & Wang, Y. Facilitation and depression at single central synapses. *Neuron* **14**, 795–802 (1995).
20. Tong, G. & Jahr, C. E. Multivesicular release from excitatory synapses of cultured hippocampal neurons. *Neuron* **12**, 51–59 (1994).
21. Debanne, D., Guerineau, N. C., Gahwiler, B. H. & Thompson, S. M. Paired-pulse facilitation and depression at unitary synapses in rat hippocampus: quantal fluctuation affects subsequent release. *J. Physiol.* **491**, 163–176 (1996).
22. Edwards, F. A. LTP is a long-term problem. *Nature* **350**, 271–272 (1991).
23. Bekkers, J. M. & Stevens, C. F. NMDA and non-NMDA receptors are colocalized at individual excitatory synapses in cultured rat hippocampus. *Nature* **341**, 230–233 (1989).
24. Liu, G. & Tsien, R. W. Properties of synaptic transmission at single hippocampal synaptic boutons. *Nature* **476**, 404–408 (1995).
25. Murthy, V. N., Sejnowski, T. J. & Stevens, C. F. Heterogeneous release properties of visualized individual hippocampal synapses. *Neuron* **18**, 599–612 (1997).

Acknowledgements. This work was supported by grants from the NIH (S.J.S.) and the Swiss National Science Foundation (H.R.). We thank W. Hopf for neuronal cultures used in some experiments.

Correspondence and requests for materials should be addressed to T.A.R. at Cornell (e-mail: taryan@mail.med.cornell.edu).

Cloning and characterization of a mammalian proton-coupled metal-ion transporter

Hiromi Gunshin*, Bryan Mackenzie*, Urs V. Berger*, Yoshimi Gunshin*, Michael F. Romero†, Walter F. Boron‡, Stephan Nussberger*, John L. Gollan‡ & Matthias A. Hediger*§

* Renal Division, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA

† Department of Cellular & Molecular Physiology, Yale University, New Haven, Connecticut 06510, USA

‡ Gastroenterology Division, Department of Medicine, Brigham & Women's Hospital, and § Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

Metal ions are essential cofactors for a wealth of biological processes, including oxidative phosphorylation, gene regulation and free-radical homeostasis. Failure to maintain appropriate levels of metal ions in humans is a feature of hereditary haemochromatosis¹, disorders of metal-ion deficiency, and certain neurodegenerative diseases². Despite their pivotal physiological roles, however, there is no molecular information on how metal ions are actively absorbed by mammalian cells. We have now identified a new metal-ion transporter in the rat, DCT1, which has an unusually broad substrate range that includes Fe²⁺, Zn²⁺, Mn²⁺, Co²⁺, Cd²⁺, Cu²⁺, Ni²⁺ and Pb²⁺. DCT1 mediates active transport that is proton-coupled and depends on the cell membrane potential. It is a 561-amino-acid protein with 12 putative membrane-spanning domains and is ubiquitously expressed, most notably in the proximal duodenum. DCT1 is upregulated by dietary iron deficiency, and may represent a key mediator of intestinal iron absorption. DCT1 is a member of the 'natural-resistance-associated macrophage protein' (Nramp) family^{3–5} and thus its properties provide insight into how these proteins confer resistance to pathogens.

In mammals, iron is thought to be transported into cells by transferrin-receptor-mediated endocytosis. However, since the available apo-transferrin in the intestinal lumen (that arising from biliary excretion) is insufficient to account fully for dietary iron

absorption⁶, it is likely that other non-receptor-mediated uptake systems exist in the intestine. Dietary iron comprises two forms, haem and non-haem iron. The bulk of intestinal non-haem iron is absorbed in the first portion of the duodenum, in which the acidic environment promotes solubilization of iron rendered in its 2+ reduced state by ferrireductase^{7,8} and ascorbate^{9,10}. How Fe²⁺ is subsequently absorbed however is poorly understood. The divalent cation transporters identified so far include the mono- or oligo-specific transporters derived from plants and yeast¹¹. In addition, mammalian divalent cation export systems have been described for zinc¹² and copper¹³, but a mammalian uptake system has not yet been identified.

We report the expression cloning, tissue distribution and initial characterization of a divalent-cation transporter, DCT1. By screening for iron uptake activity in oocytes, we isolated the DCT1 complementary DNA from a cDNA library prepared using duodenal messenger RNA from rats fed on a low-iron diet (Fe(-)D). *Xenopus* oocytes injected with Fe(-)D mRNA showed a 7-fold increase in ⁵⁵Fe uptake compared with water-injected (control) oocytes (Fig. 1). Following size-fractionation, maximal ⁵⁵Fe uptake activity was induced by a 4.0–4.5-kilobase (kb) mRNA fraction, from which we isolated a single cDNA (DCT1) which stimulated ⁵⁵Fe²⁺ uptake 200-fold (Fig. 1). The 4,409-base-pair (bp) DCT1 cDNA encodes a 561-amino-acid protein with 12 putative membrane-spanning domains (Fig. 2), predicted glycosylation sites in the fourth extracellular loop, and a consensus transport motif in the fourth intracellular loop^{3–5}. In bacteria, this motif is believed to participate in the interaction with ATP-coupling subunits; however its function is unknown and it is distinct from the nucleotide-binding fold of ABC transporters.

High-stringency northern-blot analysis of DCT1 transcripts (Fig. 3a) revealed prominent bands of 4.5 kb from proximal intestine, kidney, thymus and brain, and fainter bands in testis, liver, colon, heart, spleen, skeletal muscle, lung, bone marrow and stomach. By loading an equal amount of each mRNA, we found that DCT1 was expressed at a much higher level in proximal intestine than in

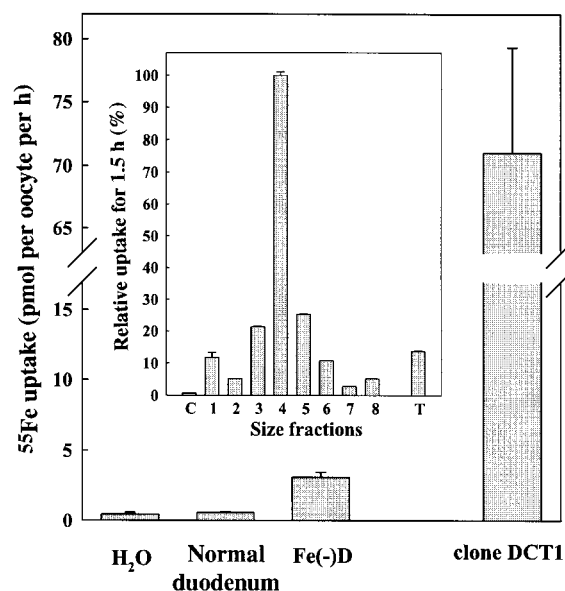


Figure 1 Uptake of 10 μ M ⁵⁵Fe in *Xenopus* oocytes injected with poly(A)⁺ RNA from normal or iron-deficient (Fe(-)D) rat duodenum or with RNA synthesized from DCT1 cDNA. Data are mean $\pm$ s.e.m. from 6–10 oocytes. Inset, relative ⁵⁵Fe uptake in oocytes injected with size-fractionated poly(A)⁺ Fe(-)D RNA. C, Control water-injected oocytes; 1, 2.0–3.0 kb poly(A)⁺ RNA; 2, 2.5–3.5 kb; 3, 3.3–4.0 kb; 4, 3.8–4.5 kb; 5, 4.4–5.7 kb; 6, 4.5–6.0 kb; 7, 4.8–6.4 kb; 8, 5.5–7.0 kb; and T, unfractionated poly(A)⁺ Fe(-)D RNA.

kidney, and more so in kidney than in brain. We observed another strong band at ~3.5 kb in kidney and thymus, indicating that a second isoform may exist in those tissues. In the intestine, DCT1 mRNA expression was highest in duodenum, and decreased towards the colon (Fig. 3b, top row). Following diet-induced iron deficiency, DCT1 mRNA expression was enhanced dramatically in duodenum (Fig. 3b, bottom row) and also, but to a lesser extent, in kidney, liver, brain, heart, lung and testis (not shown), suggesting that there is marked regulation of DCT1 mRNA by dietary iron and/or tissue iron concentration.

We investigated the cellular localization of DCT1 mRNA in small intestine, kidney, testis, thymus and brain, in frozen sections using non-radioactive *in situ* hybridization with a digoxigenin-labelled cRNA probe. High-stringency hybridization yielded specific signals with antisense probe, but no labelling with sense probe (Fig. 4). In small intestine, DCT1 was highly expressed in enterocytes lining the villus, especially in the crypts and lower segments of the villi, but not

at villar tips (Fig. 4a). Again, a proximal-to-distal gradient of expression was evident in the small intestine. This pattern of expression is consistent with the primary sites that are responsible for intestinal absorption of most divalent cations⁹.

In kidney, DCT1 mRNA labelling was most prominent in S3 proximal tubule segments (Fig. 4b), suggesting that it is involved in the reabsorption of divalent cations. Staining was also significant over the entire length of the collecting ducts, where DCT1 may participate in the final reabsorption of these ions. In testis, DCT1 mRNA was expressed in the Sertoli cells of seminiferous tubules (Fig. 4f), and was more abundant in those tubules containing mature spermatocytes. In thymus, DCT1 labelling was positive in cortical, but not medullary, thymocytes (Fig. 4d).

In each region examined in brain, DCT1 mRNA was found in neurons but not glial or ependymal cells (Fig. 4e). A qualitative examination of sagittal sections indicated that most neurons expressed DCT1 mRNA at low levels. More prominent labelling

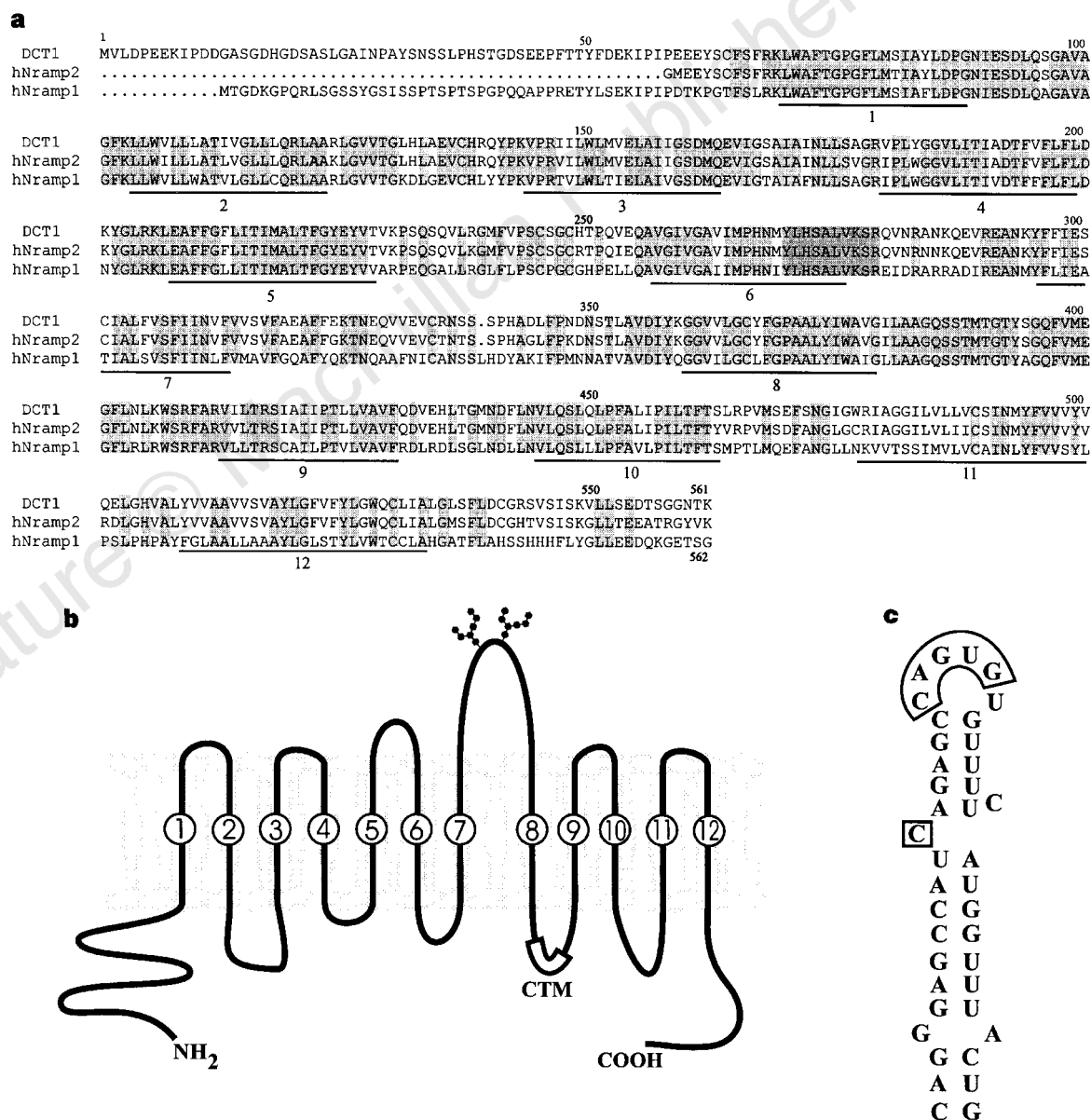


Figure 2 a, Sequence alignment of DCT1 and Nramp-related polypeptides. The sequences of rat DCT1, human Nramp1 and Nramp2 (GenBank accession numbers, L32185 and L37347) were aligned using the GCG program, and identical residues are indicated by shading. The 12 putative transmembrane regions are underlined and numbered 1-12. **b**, Twelve-transmembrane-domain model of the

DCT1 protein. Putative transmembrane domains 1 to 12 are indicated. The 'consensus transport motif'²³⁻⁵ is indicated (CTM) in the fourth intracellular loop, and putative *N*-linked glycosylation sites are identified in the fourth extracellular loop. **c**, The potential iron-responsive regulatory-protein-binding site (IRE) consensus sequence (CNCNNNCAGUG)^{24,25}, predicted to form a stem-loop.

was present in densely packed cell groups, such as the hippocampal pyramidal and granule cells (Fig. 4g), cerebellar granule cells, the preoptic nucleus and pyramidal cells of the piriform cortex. Moderate amounts of DCT1 mRNA were present in the substantia nigra (Fig. 4j). In Parkinson's disease, iron accumulates in affected neurons of the substantia nigra²; this increased iron deposition may contribute to neuron death by inducing the production of hydroxyl radicals according to the Fenton reaction. One cell group, the ventral portion of the anterior olfactory nucleus (Fig. 4i), contained large amounts of DCT1 mRNA, at least an order of magnitude more than elsewhere in the brain. DCT1 mRNA was localized in epithelial cells of the choroid plexus (Fig. 4k). Although the different functions of DCT1 in the central nervous system still need to be investigated, it is likely that defects in DCT1 contribute to the aetiology of certain neurodegenerative diseases by promoting the generation of reactive oxygen species by divalent cations.

We investigated the functional characteristics of DCT1 by using oocytes injected with DCT1 mRNA. Two-microelectrode voltage-clamp analysis indicated that divalent-cation transport mediated by DCT1 was rheogenic, with currents of up to $-1,000$ nA (Fig. 5). Whereas Fe^{2+} evoked no appreciable current (<5 nA) in control oocytes clamped at -50 mV (Fig. 5a (1)), in oocytes expressing DCT1 (Fig. 5a (2)) we observed large inward currents (I^{Fe}) when $50 \mu\text{M}$ Fe^{2+} was superfused at pH 5.5. In DCT1-injected oocytes in the absence of substrate, switching extracellular pH (pH_o) from 7.5 to 5.5 caused a larger shift in the baseline (an inward current) than in control oocytes. In two oocytes expressing DCT1, this pH-dependent current was -24.0 nA and -23.5 nA (each the average of several replicates), compared with -5.7 ± 2.7 nA (mean $\pm$ s.d.) in control oocytes from the same batch. Ca^{2+} did not mimic the effect of Fe^{2+} ; instead, 10 mM Ca^{2+} produced a small outward current at pH_o 5.5 and partially inhibited the current evoked by $50 \mu\text{M}$ Fe^{2+} (Fig. 5a (2)). Large inward currents were obtained when Zn^{2+} , Mn^{2+} , Cu^{2+} , Co^{2+} or Cd^{2+} were applied (each at $50 \mu\text{M}$); smaller currents were evoked by Ni^{2+} and Pb^{2+} (Fig. 5b).

Following step-changes in membrane potential (V_m) of oocytes expressing DCT1, we observed presteady-state and steady-state currents at pH_o 5.5 (Fig. 5c). These comprised a fast capacitive transient (decaying with a time constant of ~ 1 ms), which was also observed in control oocytes, and a transporter-mediated presteady-state current (half-time, $\tau = 30$ – 70 ms), decaying to the steady-state value (Fig. 5c). DCT1-mediated presteady-state currents were abolished in the presence of divalent cations serving as substrates (Fig. 5b) or 10 mM Ca^{2+} , and were H^+ -dependent, suggesting that H^+ may bind to the transporter—our first indication that DCT1-mediated metal-ion transport might be H^+ -coupled. The presteady-state currents were similar to the charge movements reported for other H^+ -coupled transporters^{14–16}, attributed to reorientation of a charged carrier or to H^+ binding or dissociation in the plane of the

membrane. DCT1-mediated presteady-state currents decayed more slowly than those of other cloned ion-coupled transporters (see Table 1 in ref. 15) except for the $\text{Na}^+/\text{Cl}^-/\text{GABA}$ transporter¹⁷; the V_m midpoint of charge transfer ($V_{0.5}$) at pH_o 5.5 was $>+70$ mV.

Steady-state Fe^{2+} -evoked currents were saturable (Fig. 5d): the Fe^{2+} concentration at which current was half-maximal ($K_{0.5}^{\text{Fe}}$) was $2 \mu\text{M}$ and did not vary with V_m . The maximal Fe^{2+} -evoked current ($I_{\text{max}}^{\text{Fe}}$) showed a curvilinear dependence on V_m (Fig. 5e), approaching a zero-current asymptote at positive V_m . $I_{\text{max}}^{\text{Fe}}$ was approximately linear at hyperpolarized V_m and did not saturate with hyperpolarization within the V_m range tested. The kinetic characteristics of DCT1-mediated Fe^{2+} transport were confirmed in radiotracer studies (not shown). At pH_o 6.2, ^{55}Fe uptake in NaCl medium was 88 ± 12 pmol over 1.5 h per oocyte ($n = 8$ oocytes), compared with 100 ± 10 , 118 ± 12 and 118 ± 16 pmol per 1.5 h per oocyte respectively in NaNO_3 ($n = 9$), NaSCN ($n = 6$), and choline chloride ($n = 6$). Therefore, DCT1-mediated transport does not additionally depend on extracellular Na^+ or Cl^- .

We further investigated the possibility of H^+ -coupling of metal-ion transport by simultaneously monitoring intracellular pH (pH_i) changes and Fe^{2+} -evoked currents (Fig. 5f). In an oocyte superfused with pH_o 7.5 medium at $V_h = -90$ mV, pH_i was ~ 7.35 and remained stable. The inward shift in baseline current after switching to pH_o 5.5 (in the absence of metal ion substrate) was associated with a significant decrease in pH_i that was not seen in control oocytes. Applying $50 \mu\text{M}$ Fe^{2+} at pH_o 5.5 induced a much faster and larger intracellular acidification, together with an inward current that peaked at about -200 nA. This manoeuvre had no effect in control oocytes. Thus DCT1-mediated Fe^{2+} transport is H^+ -coupled. Removal of Fe^{2+} halted the decline in pH_i and I^{Fe} returned to baseline.

At any given V_m , the current evoked by $10 \mu\text{M}$ Fe^{2+} was smaller at higher pH_o , so that the I^{Fe}/V_m relationship appeared to be shifted to the left with increasing pH_o (Fig. 5g): that is, Fe^{2+} transport was driven at higher rates at hyperpolarized potentials and/or low pH_o . I^{Fe} increased in magnitude as a hyperbolic function of $[\text{H}^+]_o$ (Fig. 5h); current was half-maximal at $1.3 \mu\text{M}$ H^+ ($K_{0.5}^{\text{H}}$). The Hill coefficient for H^+ (n_{H}) was about 1 (as for iron; Fig. 5d) and did not vary appreciably from unity at each V_m tested (not shown). Hill analysis suggested a transport stoichiometry for DCT1 of $1 \text{ H}^+ : 1 \text{ Fe}^{2+}$. At physiological membrane potentials (-90 to -30 mV), the apparent affinity constant for H^+ ($K_{0.5}^{\text{H}}$) was 1 – $2 \mu\text{M}$. H^+ concentrations of at least $1 \mu\text{M}$ (pH 6.0) are expected in the brush-border extracellular microenvironment¹⁸ in the small intestine, even after the bulk luminal pH has risen substantially.

What is the physiological significance of coupling of divalent-cation transport with H^+ ? First, coupling to the movement of H^+ down its electrochemical gradient will always ensure a concentrative uptake of divalent cations, even at trace amounts. Second, it is

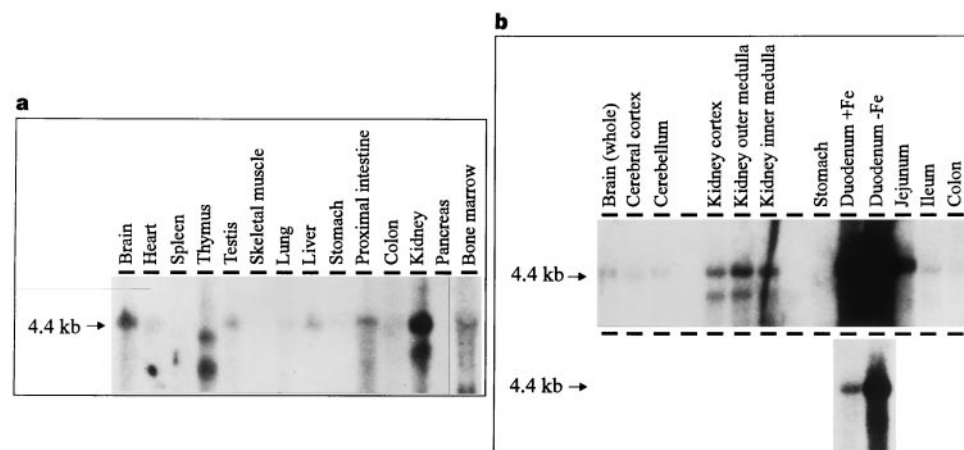


Figure 3 High-stringency northern blot analysis of RNA from rat tissues probed with ^{32}P -labelled DCT1 cDNA. Each well was loaded with $3 \mu\text{g}$ poly(A)⁺ RNA from a range of **a**, whole tissues, or **b** (top row), specific regions of brain, kidney and intestine, including duodenum from rats fed a low-iron diet ($\text{Fe}(-)\text{D}$); bottom row, duodenal samples after short exposure.

possible that cotransported H^+ could result in a slight lowering of pH at the intracellular surface of the membrane, sufficient to maintain the solubilization of iron before its binding to intracellular proteins. In addition to operating as a H^+ /divalent cation cotransporter (symporter), DCT1 can apparently operate as a H^+ uniporter: that is, a H^+ 'leak' pathway can proceed uncoupled from divalent cation transport. This was demonstrated by the inward currents and intracellular acidification observed after switching pH_o from 7.5 to 5.5 and is analogous to the Na^+ /glucose cotransporter SGLT1 in which the phlorizin-inhibitable leak current in the absence of sugar is carried by Na^+ (ref. 19). Similar observations have been made for other ion-coupled transporters, such as the plant H^+ /amino acid²⁰ and H^+ /sucrose¹⁶ cotransporters, and a low-affinity homologue of SGLT1 (ref. 21). It is unlikely that these phenomena were due to H^+ -coupled transport of Ca^{2+} or Mg^{2+} (present at low concentrations in the medium) because, instead of increasing the inward currents, addition of excess (5 mM) Mg^{2+} had no effect, and excess Ca^{2+} (which partially inhibited the Fe^{2+} -evoked current) resulted in a small outward current, presumably as a result of blocking the H^+ leak pathway. The effects of Ca^{2+} (which, like Fe^{2+} , also abolished the presteady-state currents) were reminiscent of the effects of phlorizin on the Na^+ /glucose cotransporters^{19,21}. Our data therefore suggest that Ca^{2+} is a blocker (or very weak agonist), reacting with DCT1 at very low affinity ($K_i \approx 10$ mM).

Further kinetic measurements will distinguish between co-transporter models²², by showing for example which ligand (H^+ or divalent cation) binds first and whether they are transported

simultaneously or consecutively. The biophysical properties of the DCT1 transporter are strikingly similar to those of several H^+ - or Na^+ -coupled transporters (see Table 1 in ref. 15)¹⁴ that serve a wide range of substrates (including sugars, oligopeptides, amino acids), despite a lack of significant sequence homology.

Taken together, our data show that there is an active, cellular uptake mechanism for divalent cations (including Fe^{2+}) in mammalian cells. DCT1 accepts a broad range of metal ions, favouring the divalent cations Fe^{2+} , Cd^{2+} , Co^{2+} , Cu^{2+} , Ni^{2+} , Mn^{2+} , Pb^{2+} and Zn^{2+} among those tested. The existence of a common intestinal absorptive mechanism for a range of metals has important nutritional implications. Absorption of these divalent cations will proceed even at physiological Ca^{2+} levels. However our results suggest that excessive luminal Ca^{2+} could interfere with the normal absorption of trace metals (see also ref. 23). DCT1 also transports the toxic heavy metals Cd^{2+} and Pb^{2+} . We cannot at present exclude the involvement of DCT1 in receptor-mediated endocytic absorption of iron (endosomal pH of about 6.0), a possibility that could be tested using immunocytochemistry.

DCT1 was upregulated in response to dietary iron deficiency in most, if not all, tissues. DCT1 cDNA contains in its 3' untranslated region a putative iron-responsive element which forms a stem-loop containing the consensus sequence (Fig. 1c) found in the 3' and 5' iron-responsive elements of the transferrin receptor and the ferritin mRNAs, respectively^{24,25}. By analogy with transferrin receptor mRNA, the iron-responsive element may regulate DCT1 mRNA levels by RNA degradation mediated by a protein that binds to this element.

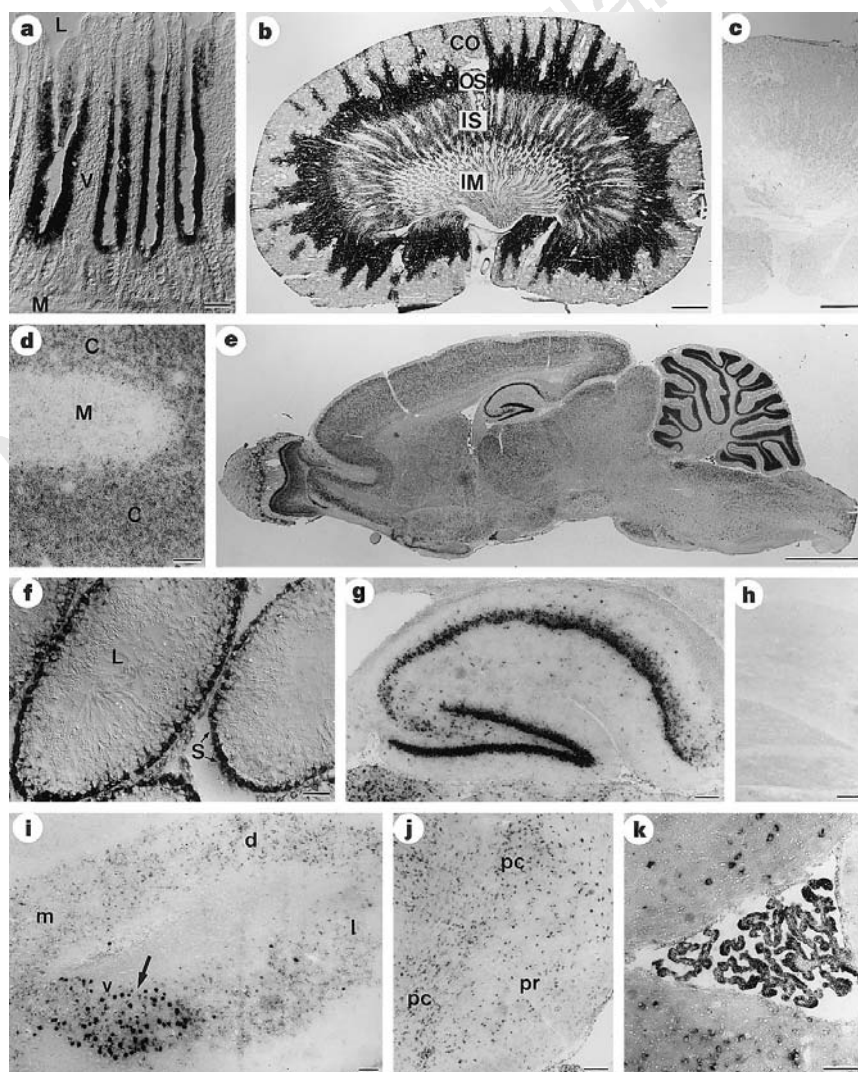


Figure 4 Tissue localization of rat DCT1 mRNA detected by *in situ* hybridization. Bright-field micrographs of cryosections hybridized to digoxigenin-labelled DCT1 antisense (or sense in **c**, **h**) cRNA probe. **a**, Duodenum (M, muscle layer; V, villi; L, lumen). **b**, Kidney (CO, cortex; OS, outer stripe of the medulla; IS, inner stripe; IM, inner medulla). **c**, Kidney hybridized with sense probe. **d**, Thymus (C, cortex; M, medulla). **e**, Sagittal section through brain; DCT1 mRNA was found in neurons throughout the brain. **f**, Testis (S, Sertoli cells; L, lumen). **g**, In hippocampus, positive labelling was prominent in the pyramidal and granule cells. No labelling was obtained with sense probe (**h**). **i**, Anterior olfactory nucleus (m, medial; d, dorsal; l, lateral; v, ventral). **j**, Substantia nigra (pr, pars reticulata; pc, pars compacta). **k**, Choroid plexus in the fourth ventricle. Scale bars: 100 μ m in **a**, **d**, **f**, **i**, **k**; 1.5 mm in **b**, **c**; 3 mm in **e**; and 200 μ m in **g**, **h**, **j**.

The marked responsiveness of DCT1 expression (at least at the mRNA level) to manipulations of dietary iron has implications for hereditary haemochromatosis (HH). This common autosomal recessive disease occurs primarily in individuals of northern European origin. The primary pathophysiological defect is in the small intestine where the absorption of iron in HH patients is increased, leading to progressive iron deposition in the liver, heart and other organs. Interestingly, HH patients also accumulate other metals, such as Co^{2+} and Mn^{2+} (ref. 26). As this is consistent with the broad substrate range of DCT1, we propose that DCT1 is constitutively upregulated as an indirect result of a defect in the recently identified haemochromatosis gene *HFE*. A single point mutation in *HFE* (resulting in the amino-acid substitution Cys 282 $\rightarrow$ Tyr) is found in the majority of HH patients¹. As the *HFE* gene product corresponds to a protein related to the major histocompatibility complex, with only a single membrane-spanning region and a short

cytoplasmic tail, it is unlikely that HFE is a transporter. Instead, HFE may be involved in 'sensing' iron levels and regulating the expression of other gene products, including DCT1, thereby explaining how a defective *HFE* gene product might result in iron overload.

DCT1 is the most recently identified member of an emerging family (Nramp) of mammalian proteins (Fig. 2a). Among these, the mouse macrophage protein Nramp1 confers resistance to mycobacterial infection in mice, but the host defence mechanism and biochemical function of Nramp1 are unknown⁴. The amino-acid sequence of DCT1 is 73% identical to human Nramp1 and probably corresponds to the rat isoform of the ubiquitously expressed human Nramp2 (with which it shares 92% identity). Based on our functional characterization of DCT1 expressed in oocytes, we propose that the Nramp family of proteins are divalent-cation transporters. We found that ^{55}Fe uptake was stimulated in oocytes expressing

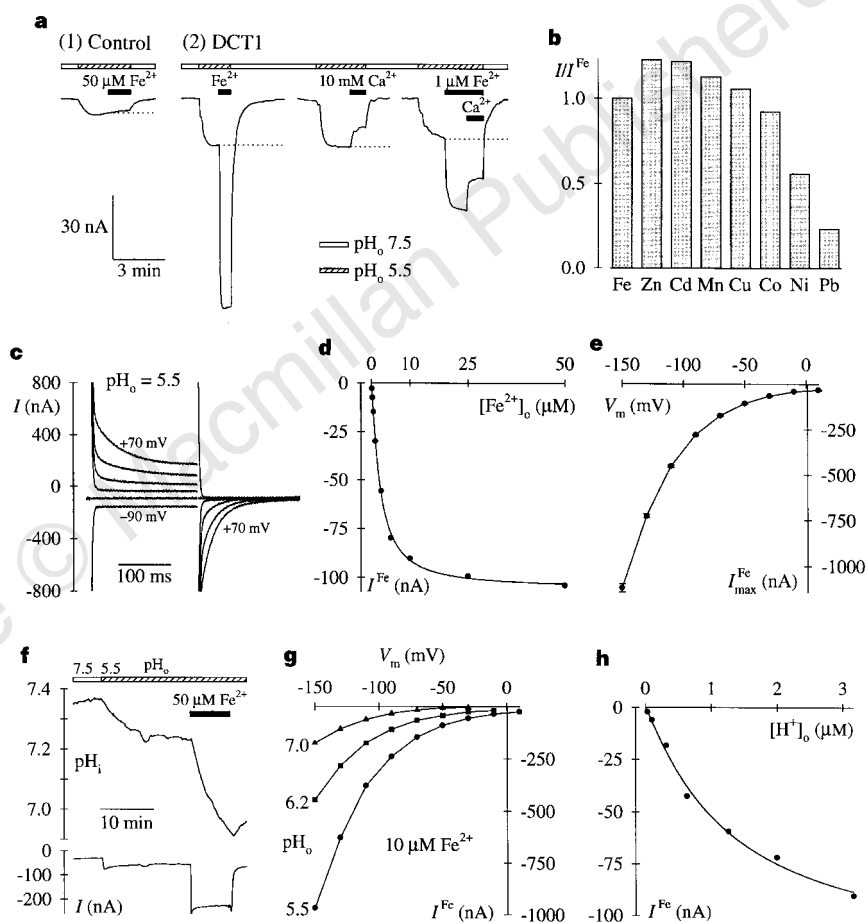


Figure 5 Currents associated with the divalent cation transporter DCT1 expressed in oocytes. **a**, Current was continuously monitored in (1) a control oocyte and (2) a single oocyte expressing DCT1, each clamped at -50 mV and superfused for periods indicated by the boxes in the top panel, first at $\text{pH } 7.5$ (blank) then at $\text{pH } 5.5$ (diagonal hatch). Iron or additional calcium was applied for the periods shown by the solid bars, then washed out with substrate-free medium, $\text{pH } 7.5$. The dotted horizontal lines indicate the approximate baseline obtained in $\text{pH } 5.5$ medium before the addition of test substrates. **b**, Substrate specificity of DCT1, applying test substrates at $50 \mu\text{M}$ ($\text{pH}_o 5.5$) in a single oocyte ($V_h = -50 \text{ mV}$); evoked currents were normalized to the Fe^{2+} -evoked current (-98.5 nA). **c**, An oocyte expressing DCT1 (superfused at $\text{pH}_o 5.5$) was held at V_h of -50 mV , and V_m was stepped to between $+90 \text{ mV}$ and -110 mV (in 20 mV increments) for 200 ms . Currents are displayed from 3 ms after the voltage step (on and off); for clarity, only the currents obtained at V_m of $+70$, $+50$, $+30$, -10 , -50 and -90 mV are shown. **d**, Concentration-dependence of the

Fe^{2+} -evoked currents ($\text{pH}_o 5.5$, $V_h = -50 \text{ mV}$). The currents evoked by 0.1 , 0.25 , 0.5 , 1 , 2.5 , 5 , 10 , 25 and $50 \mu\text{M FeCl}_2$ were fitted to equation (1): $I_{\text{max}}^{\text{Fe}} = -105 \pm 1 \text{ nA}$, $K_{0.5}^{\text{Fe}} = 2.2 \pm 0.1 \mu\text{M}$, and $n_H = 1.2 \pm 0.1$ for iron. **e**, Voltage-dependence of $I_{\text{max}}^{\text{Fe}}$ (neither $K_{0.5}^{\text{Fe}}$ nor n_H varied with V_m). **f**, Simultaneous recordings of intracellular pH (pH_i) changes and current in an oocyte expressing DCT1. The oocyte was clamped at $V_h = -90 \text{ mV}$ and superfused for the periods indicated by the boxes in the top panel, with $\text{pH } 7.5$ medium (blank), then with $\text{pH } 5.5$ medium (diagonal hatch). $50 \mu\text{M FeCl}_2$ was added for the period shown by the solid bar. **g**, Currents evoked by $10 \mu\text{M Fe}^{2+}$ as a function of extracellular pH (pH_o). **h**, Proton activation of the currents evoked by $10 \mu\text{M Fe}^{2+}$ at $V_h = -50 \text{ mV}$. Data were fitted to equation (1): $I_{\text{max}}^{\text{H}} = -121 \pm 21 \text{ nA}$ (compare with $I_{\text{max}}^{\text{Fe}}$), $K_{0.5}^{\text{H}} = 1.3 \pm 0.5 \mu\text{M}$, $n_H = 1.1 \pm 0.2$ for H^+ . ($I_{\text{max}}^{\text{H}}$ varied with V_m in a similar manner to $I_{\text{max}}^{\text{Fe}}$.) All kinetic data presented (**d**, **e**, **g**, **h**) were obtained from the same oocyte; errors represent the error in the estimates of kinetic parameters (equation (1), see Methods).

mouse Nramp1 (not shown), although the apparent affinity for Fe^{2+} was much lower than for DCT1. Functioning as a divalent-cation transporter, Nramp1 could confer resistance to infection by either of the following mechanisms. (1) H^+ -coupled uptake of Fe^{2+} in macrophages by Nramp1 may provide these cells with Fe^{2+} to produce toxic hydroxyl radicals via the Fenton reaction, killing pathogens in phagosomes as part of the defence mechanism. (2) Fe^{2+} , Mn^{2+} or other divalent cations may be depleted from the phagosome by Nramp1 in the phagosomal membrane. This could result in a lack of essential metal ions which are required by the pathogen for the activity of enzymes such as superoxide dismutase and catalase in order to detoxify superoxide anions or hydroxyl radicals, thereby preventing propagation of the bacteria. A similar proposal²⁷ has been made for the SMF1 transporter, a yeast homologue of the Nramp family.

Collectively, the data presented constitute a major advance in our understanding of a critical biological question: how divalent cations cross cell membranes, particularly those of epithelia. Moreover, the existence of a single transport mechanism serving a range of divalent cations and driven by the proton electrochemical gradient has nutritional, clinical and toxicological implications. □

Methods

Expression cloning. DCT1 cDNA was cloned from rats fed on a low-iron diet (ICN) with milli-Q water *ad libitum* (3–4 weeks). Briefly, RNA was extracted from duodenal mucosal scrapes from iron-deficient ($\text{Fe}(-)\text{D}$) male rats and size-fractionated using preparative gel electrophoresis. A positive (~4.5 kb) poly(A)⁺ RNA fraction, which stimulated ^{55}Fe uptake in oocytes, was used to construct a directional cDNA library using a SuperScript cDNA synthesis system (GibcoBRL). cDNA was ligated into the pSPORT1 vector and ElectroMax DH10B cells were transformed as described²⁸. RNA synthesized *in vitro* from pools of ~200 clones was injected into oocytes. A positive pool was sequentially subdivided and analysed until we identified a single clone (DCT1) that was sequenced at the Yale University Sequencing Facility.

Radiotracer uptake. Collagenase-treated *Xenopus* oocytes were injected with ~50 ng poly(A)⁺ RNA or ~25 mg DCT1 RNA synthesized *in vitro*. Radiotracer uptake was determined (3 d after RNA injection) by incubating 6–10 oocytes for 1 or 1.5 h in 750 μl standard uptake medium (100 mM NaCl, 10 mM HEPES, 1 mM ascorbic acid, pH 6.2) together with 10 μM $^{55}\text{FeCl}_2$. Oocytes were solubilized using 10% SDS and the ^{55}Fe content was measured by liquid scintillation counting.

Northern analysis. Poly(A)⁺ RNA (3 μg each well) was separated on a 18% formaldehyde/1% agarose gel and blotted onto a nitrocellulose filter (Schleicher and Schuell). The DCT1 insert was excised from pSPORT1 and labelled with ^{32}P using a T7 QuickPrime kit (Pharmacia). Hybridization was for 16 h at 42 °C in 50% formamide, and filters were washed in $5 \times \text{SSC}/0.1\%$ SDS at 50 °C for 2×30 min, $0.1 \times \text{SSC}/0.1\%$ at 65 °C for 3×20 min.

In situ hybridization. Digoxigenin-labelled antisense and sense run-off transcripts were synthesized using a Genius Kit (Boehringer-Mannheim) from a PCR fragment, which contained about 1.7 kb of DCT1 sequence (bases 105–1,788) and which was flanked by promoter sites for SP6 and T7 RNA polymerase. Transcripts were alkali-hydrolysed to an average length of 200–400 base pairs. *In situ* hybridization was performed on cryosections (12 μm) of fresh-frozen tissue as described²⁹. The hybridization buffer consisted of 50% formamide, $5 \times \text{SSC}$, 2% blocking reagent (Boehringer-Mannheim), 0.02% SDS, 0.1% *N*-laurylsarcosine; probe concentrations were ~200 ng ml^{-1} . Sections in slide mailers were immersed in hybridization solution and hybridized at 70 °C for 18 h. Sections were then washed 3 times in $2 \times \text{SSC}$ and for 2×30 min in $0.1 \times \text{SSC}$ at 70 °C. The hybridized labelled probes were visualized using anti-digoxigenin Fab fragments (Boehringer-Mannheim) and BCIP/NBT substrate²⁹. Sections were developed in substrate solution for 42 h, then rinsed in 10 mM Tris, 1 mM EDTA, pH 8.0, and coverslipped with 50% PBS/glycerol.

Electrophysiology. A two-microelectrode voltage clamp was used to measure steady-state and presteady-state currents in control oocytes and oocytes injected with ~25 ng of *in vitro* transcribed DCT1 RNA, 4–7 days after injection. Microelectrodes (resistance 0.5–5 M Ω) were filled with 3 M KCl.

Oocytes were superfused at 23–24 °C in a low-calcium medium with ascorbic acid (96 mM NaCl, 2 mM KCl, 0.6 mM CaCl_2 , 1 mM MgCl_2 , 100 μM ascorbic acid, 2.5 mM HEPES, 2.5 mM MES, buffered to between 5.5 and 7.5 with Tris) and clamped at a holding potential (V_h) of –50 mV. The current was low-pass filtered at 5 kHz and digitized at 5 kHz. Step-changes in membrane potential (V_m) were applied (from +50 mV to –150 mV for steady-state studies, and +90 mV to –110 mV for presteady-state analysis, in 20-mV increments), each for a duration of 200 ms, at pH_o 5.5–7.5 before and after the addition of the test substrate. Additionally (Fig. 5a and b), current was monitored continuously (sampling at 1 Hz) in oocytes clamped at –50 mV, but without step changes in V_m . Test solutions were washed out by superfusing the oocyte with substrate-free medium at pH 7.5 for several minutes. Steady-state data (obtained by averaging the points over the final 11 ms at each V_m) were fitted to equation (1), for which I is the evoked current (that is, the difference in steady-state current measured in the presence and absence of substrate), I_{max} the derived current maximum, S the substrate (divalent cation or H^+) concentration, $K_{0.5}$ the substrate concentration at which current was half-maximal, and n_H the Hill coefficient

$$I = \frac{I_{\text{max}} S^{n_H}}{(K_{0.5})^{n_H} + S^{n_H}} \quad (1)$$

Intracellular pH measurement. Changes in intracellular pH (pH_i) associated with DCT1-mediated transport in oocytes were measured using pH microelectrodes. We used silanized borosilicate pipettes back-filled with phosphate buffer at pH 7.0, with the tips filled with the hydrogen ionophore I-cocktail B (Fluka) as described³⁰. pH_i and current (filtered at 25 Hz) were measured simultaneously and digitized at 0.5 Hz in voltage-clamped oocytes, according to a method to be described elsewhere (M.F.R. and W.F.B., manuscript in preparation).

Received 8 April; accepted 19 June 1997.

- Feder, J. N. *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genet.* **13**, 399–408 (1996).
- Hirsch, E. C. Biochemistry of Parkinson's disease with special reference to the dopaminergic systems. *Mol. Neurol.* **94**, 135–142 (1997).
- Gruenheid, S., Cellier, M., Vidal, S. & Gros, P. Identification and characterization of a second mouse *Nramp* gene. *Genomics* **25**, 514–525 (1995).
- Vidal, S., Malo, D., Vogan, K., Skamene, E. & Gros, P. Natural resistance to infection with intracellular parasites: isolation of a candidate for *Bcg*. *Cell* **73**, 469–485 (1993).
- Vidal, S., Gros, P. & Skamene, E. Natural resistance to infection with intracellular parasites: molecular genetics identifies *Nramp1* as the *Bcg/Ity/Lsh* locus. *J. Leuk. Biol.* **58**, 382–390 (1995).
- Conrad, M. E., Umbreit, J. N., Moore, E. G. & Heiman, D. Mobilferrin is an intermediate in iron transport between transferrin and hemoglobin in K562 cells. *J. Clin. Invest.* **98**, 1449–1454 (1996).
- Raja, K. B., Simpson, R. J. & Peters, T. J. Investigation of a role for reduction in ferric iron uptake by mouse duodenum. *Biochim. Biophys. Acta* **1135**, 141–146 (1992).
- Jordan, I. & Kaplan, J. The mammalian transferrin-independent iron transport system may involve a surface ferrireductase activity. *Biochem. J.* **302**, 875–879 (1994).
- Han, O., Failla, M. L., Hill, A. D., Morris, E. R. & Smith, J. C. Reduction of Fe(III) is required for uptake of nonheme iron by Caco-2 cells. *J. Nutr.* **125**, 1291–1299 (1995).
- Dorey, C. *et al.* Iron speciation at physiological pH in media containing ascorbate and oxygen. *Br. J. Nutr.* **70**, 157–169 (1993).
- Eide, D. J. Molecular biology of iron and zinc uptake in eukaryotes. *Curr. Opin. Cell Biol.* (in the press).
- Palmiter, R. D., Cole, T. B., Quaipe, C. J. & Findley, S. D. ZnT-3, a putative transporter of zinc into synaptic vesicles. *Proc. Natl Acad. Sci. USA* **93**, 14934–14939 (1996).
- Bull, P. & Cox, D. W. Wilson disease and Menkes disease: new handles on heavy-metal transport. *Trends Genet.* **10**, 246–252 (1994).
- Nussberger, S. *et al.* Symmetry of H^+ binding to the intra- and extracellular side of the H^+ -coupled oligopeptide cotransporter PepT1. *J. Biol. Chem.* **272**, 7777–7785 (1997).
- Mackenzie, B. *et al.* Mechanisms of the human intestinal H^+ -coupled oligopeptide transporter hPEPT1. *J. Biol. Chem.* **271**, 5430–5437 (1996).
- Boorer, K. J., Loo, D. D. F. & Wright, E. M. Steady-state and presteady-state kinetics of the H^+ /hexose cotransporter (STP1) from *Arabidopsis thaliana* expressed in *Xenopus* oocytes. *J. Biol. Chem.* **269**, 20417–20424 (1994).
- Mager, S. *et al.* Steady states, charge movements, and rates for a cloned GABA transporter expressed in *Xenopus* oocytes. *Neuron* **10**, 177–188 (1993).
- McEwan, G. T. A., Daniel, H., Fett, C., Burgess, M. N. & Lucas, M. L. The effect of *Escherichia coli* STa enterotoxin and other secretagogues on mucosal surface pH of rat small intestine *in vivo*. *Proc. R. Soc. Lond.* **234**, 219–237 (1988).
- Mackenzie, B., Loo, D. D. F. & Wright, E. M. Coupling stoichiometry for the Na^+ /glucose cotransporter SGLT1 (manuscript in preparation).
- Boorer, K. J. *et al.* Kinetics and specificity of a H^+ /amino acid transporter from *Arabidopsis thaliana*. *J. Biol. Chem.* **271**, 2213–2220 (1996).
- Mackenzie, B., Loo, D. D. F., Panayotova-Heiermann, M. & Wright, E. M. Biophysical characteristics of the pig kidney Na^+ /glucose cotransporter SGLT2 reveal a common mechanism for SGLT1 and SGLT2. *J. Biol. Chem.* **271**, 32678–32683 (1996).
- Jauch, P. & Luger, P. Electrogenic properties of the sodium–alanine cotransporter in pancreatic acinar cells: II. Comparison with transport models. *J. Membr. Biol.* **94**, 117–127 (1986).
- Gunshin, H., Noguchi, T. & Naito, H. Effect of calcium on the zinc uptake by brush border membrane vesicles isolated from the rat small intestine. *Agric. Biol. Chem.* **55**, 2813–2816 (1991).
- Casey, J. L. *et al.* Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. *Science* **240**, 924–928 (1988).

25. Klausner, R. D., Rouault, T. A. & Harford, J. B. Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell* **72**, 19–28 (1993).
26. Nichols, G. M. & Bacon, B. R. Hereditary hemochromatosis: pathogenesis and clinical features of a common disease. *Am. J. Gastroenterol.* **84**, 851–862 (1989).
27. Supek, E., Supekova, L., Nelson, H. & Nelson, N. A yeast manganese transporter related to the macrophage protein involved in conferring resistance to mycobacteria. *Proc. Natl Acad. Sci. USA* **93**, 5105–5110 (1996).
28. Hediger, M. A. & Rhoads, D. Molecular physiology of sodium–glucose cotransporters. *Physiol. Rev.* **74**, 993–1026 (1994).
29. Schaeren-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: *in situ* hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431–440 (1993).
30. Romero, M. F., Hediger, M. A., Boulpaep, E. L. & Boron, W. F. Expression cloning of the renal electrogenic Na⁺/HCO₃⁻ cotransporter. *Nature* **387**, 409–413 (1997).

Acknowledgements. Nramp cDNA was a generous gift from P. Gros. We thank D. Eide for stimulating discussion. This research was supported by the NIH.

Correspondence and requests for materials should be addressed to H.G. (e-mail: hgusshin@rics.bwh.harvard.edu) or to M.H. (Mhediger@rics.bwh.harvard.edu). The DCT1 sequence is deposited in GenBank under accession number AF008439.

Cell-free V(D)J recombination

Dale A. Ramsden, Tanya T. Paull & Martin Gellert

Laboratory of Molecular Biology, Building 5, Room 241, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0540, USA

V(D)J recombination generates diversity in the immune system through the lymphoid-specific assembly of multiple gene segments into functional immunoglobulin and T-cell receptor genes (for reviews, see refs 1, 2). The first step in V(D)J recombination is cleavage of DNA at recombination signal sequences. Cleavage produces a blunt DNA end on each signal sequence and a hairpin end on adjacent coding gene segments^{3,4}, and can be reproduced *in vitro* by using purified RAG1 and RAG2 proteins^{5,6}. The later steps involve processing and joining of the cleaved DNA ends, and until now have been studied only in cells. Here we reconstitute the complete V(D)J recombination reaction in a cell-free system. We find that the RAG proteins are not only involved in cleavage, but are also needed in the later steps for efficient joining of coding ends. Joining is largely directed by short pieces of identical sequence in the coding flanks, but addition of human DNA ligase I results in greater diversity. Coding junctions contain short deletions as well as additions complementary to a coding flank (P nucleotides). Addition of non-templated nucleotides into coding junctions is mediated by terminal deoxynucleotidyl transferase. The cell-free reaction can therefore reproduce the complete set of processing events that occur in cells.

In establishing a cell-free reaction, we concentrated on the joining of coding segments because this process involves multiple steps that are highly characteristic of V(D)J recombination. The hairpin ends from the initial cleavage are opened and then nucleotides are often removed or added before joining. The frequent retention of self-complementary inserts is also characteristic of V(D)J coding junctions^{1,2}. These P nucleotides are derived from asymmetric opening of hairpin-coding ends. By comparison, signal-end joining could be accomplished by simple ligation of two blunt DNA ends⁷.

To reconstitute complete V(D)J recombination, we used a partially fractionated HeLa cell extract as a supplement to cleavage by RAG proteins, because the factors required for the later stages are not specific to any particular cell type⁸. Using the polymerase chain reaction (PCR) (Fig. 1a), we detected coding junctions when RAG proteins, the HeLa cell fraction and high-energy cofactors were included (Fig. 1b, lanes 1–5). No joining was observed if either the RAG proteins or the cell fraction were omitted. The HeLa cell fraction was inactivated by heating (Fig. 1b, lane 3), suggesting that its activity was due to protein factors. Joining was optimal when both ATP and all four deoxynucleoside triphosphates (dNTPs) were included (Fig. 1b, lane 4), although some activity was detectable

with either ATP or dNTPs alone. In the presence of ATP, addition of dideoxynucleotides did not stimulate joining like dNTPs, suggesting that dNTPs are substrates for a DNA polymerase (data not shown). In the absence of any added nucleoside triphosphates, joining was minimal (Fig. 1b, lane 5). Coding junctions were detectable within 3 hours of addition of extract, and the amounts continued to increase for 20 h. Signal junctions were also formed under similar conditions, but did not require addition of the cell fraction if T4 DNA ligase and ATP were included (Fig. 2b).

These reactions ('standard') were done in two stages: the first included only factors required for cleavage, the second included the HeLa cell fraction and high-energy cofactors. When joining factors were included from the start, cleavage and joining were both reduced (data not shown).

To determine whether RAG proteins participate in joining as well as cleavage, we compared the standard reaction with one in which the RAG proteins were removed by deproteinization before addition of the HeLa cell fraction (Fig. 2). When RAG proteins were present in the joining phase, they stimulated formation of the coding junctions as much as 50-fold (Fig. 2a), indicating that they may direct hairpin opening and/or recruit other factors for joining coding ends.

Stimulation of coding-junction formation was not due to continued cleavage by the RAG proteins in the second stage of the reaction (confirmed by direct staining of cleaved DNA; Fig. 2c). Furthermore, a second DNA substrate (designed to give a different-sized PCR product) added only in the second stage was not recombined (with or without an additional aliquot of RAG proteins). This verifies that residual cleavage activity makes no significant contribution to the amount of joined product (data not shown).

In contrast, signal-end joining by the HeLa cell fraction was inhibited more than 5-fold in the presence of RAG proteins (Fig. 2b). This effect was specific to the HeLa cell fraction, as joining of signal ends by T4 ligase was unaffected (Fig. 2b). RAG proteins therefore promote joining of coding ends but inhibit signal joining, consistent with the observation that signal ends persist for much longer than coding ends in cells^{3,9}. Inhibition of signal joining could be due to a continued stable association of RAG proteins with signal ends¹⁰ which blocks access by mammalian joining factors.

As already noted, the processing event specific to V(D)J recombination is the frequent addition of P nucleotides, so we determined whether the junctions generated *in vitro* also have these insertions. We sequenced *in vitro* derived coding junctions and identified several examples of such P nucleotides (Table 1).

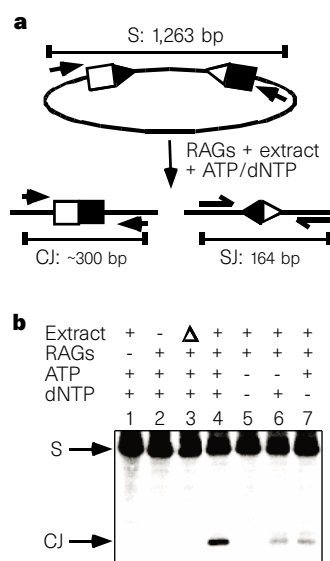


Figure 1 Requirements for cell-free V(D)J recombination. **a**, Plasmid assay. Black triangles, 12-bp spacer recombination signals; white triangles, 23-bp spacer recombination signals. PCR primers are indicated by arrows. PCR products and sizes are described above or below the primer pair used to generate them. CJ, coding joint PCR product; SJ, signal joint PCR product; S, substrate PCR product. **b**, Basic requirements for coding joint formation. RAG proteins, HeLa fraction, ATP and dNTP were included as indicated. Triangle indicates the cell fraction was heated at 68 °C for 10 min.

Coding junctions made by *V(D)J* recombination are usually highly diverse after gaining or losing nucleotides during the process². Under cell-free conditions, we observed five different junction sequences in a sample of 25 junctions (from two independent joining reactions; Table 1a). The most common junction (19/25) has one copy of a 5-base-pair (bp) direct repeat (CAGGT; direct repeats are underlined in Table 1) precisely deleted from the two coding flanks. The next most frequently observed junction (3/25) may also have a precise deletion of a copy of a 5-bp direct repeat; at this junction, the direct repeat would have been generated by extending the coding flank of the 12-spacer signal with five more P nucleotides (ACCTG). Joining directed by short homologous sequences of DNA is also a feature of *V(D)J* recombination in cells², although the length of the deletion or addition (12 bp) observed here is greater than usual. The extent to which short repeats direct joining was further characterized by insertion of 30 bp between the 5-bp identical sequence and the 12-spacer signal. In this modified substrate, most joining still deleted precisely one copy of this motif,

even though deletion of over 40 bp of coding flank sequence was now required (data not shown).

Processing of coding flanks that contain directly repeated sequences can generate complementary overhangs. The added stability provided by these overhangs may result in homology-directed joining. DNA ligases differ greatly in their ability to join ends without complementary overhangs: for example, T4 DNA ligase and mammalian DNA ligase I (ref. 11) can join blunt ends, whereas the other known mammalian DNA ligases (DNA ligases II, III and IV; refs 11–13, respectively) require complementary overhangs. (Ligase II appears to be a fragment of ligase I rather than a separate entity^{14,15}.) The preference for homology-directed joining might reflect a low ligase activity, reduced during partial fractionation of the extract or by inappropriate storage of the cell fraction (DNA ligase I, for example, loses activity after freezing¹⁶).

We therefore supplemented the HeLa cell fraction with different purified ligases. Joining under standard conditions produced a major PCR product of 283 bp, representing junctions with a 12-bp

Table 1 Structure of coding joints generated by cell-free *V(D)J* recombination

No.	12 flank CAGGTCGACTTA	(12 RSS) P	N	(23 RSS) P	23 flank CGGTGGA
(a) Extract (standard conditions)					
19	<u>CAGGT</u>				GGA
3	CAGGTCGACTTA	TAAGTCG		<u>ACCTG</u>	CAGGTGGA
1	CAGGTC				<u>GA</u>
1	CAGGTCG			<u>ACCTG</u>	CAGGTGGA
1	CAGGTCGACTTA	TAAGT			TGGA
(b) Extract + ligase I					
8	CAGGTCGACTTA				CAGGTGGA
3	<u>CAGGT</u>				GGA
2	CAGGTCGACTTA	TAAGTCG		<u>ACCTG</u>	CAGGTGGA
2	CAGGTCGACT			CCTG	CAGGTGGA
2	CAGGTCGACTTA	T		CCTG	CAGGTGGA
1	CAGGTCGACTTA	TAA		CCTG	CAGGTGGA
1	CAGGTCGA			CCTG	CAGGTGGA
1	CAGGTCGACTTA			CCTG	CAGGTGGA
1	CAGGTCGACTTA				GTGGA
1	CAGGTCGACTTA			G	AGGTGGA
1	CAGGTCGAC				CAGGTGGA
1	CAGGTCGACTTA				GGTGA
1	CAGGTCGACT				CAGGTGGA
(c) Extract + ligase I + TdT					
5	CAGGTCGATTA				CAGGTGGA
1	CAG		N ₂₇		CAGGTGGA
1	CAGGTCGACTTA	TA	GCGAC	G	CAGGTGGA
1	CAGGTCGACTT				GA
1	CAGG				GGTGA
1	CAGGTC		N ₅₄	CTG	CAGGTGGA
1	CAGGTCGACT		CGATACCGCTCTCCT		CAGGTGGA
1	CAGGTCGACT		ATCGGA	G	CAGGTGGA
1	CAGGTCGACTTA	TA			GGTGA
1	CAGGTCG		TACGAXCGGTCTCTGA		AGGTGGA
1	CAGGTCGACTT		CGGAAC	G	CAGGTGGA
1	CAGGTCGACTTA	T	GAATTGAAXC		GGTGA
1	CAGGTCGAC		CT		CAGGTGGA

Sequence flanking the 12 signal and 23 signal in the starting plasmid is shown at the top of the table. No., number of examples of a coding joint sequence. Joints at which a significant (greater than 1 bp) homology may have directed joining are underlined. P, added sequence palindromic to a coding flank; P additions are listed adjacent to their parental coding flank. N, addition of non-templated sequence (for two joints, we record only N addition length (27 and 54 bp) and not sequence). Ambiguous bases are denoted by an X. **a**, Standard conditions; **b**, reactions supplemented with DNA ligase I, as for Fig. 3a, lane 8; **c**, reactions supplemented with DNA ligase I and 1.5 U TdT (Gibco-BRL), as for Fig. 4, lane 6.

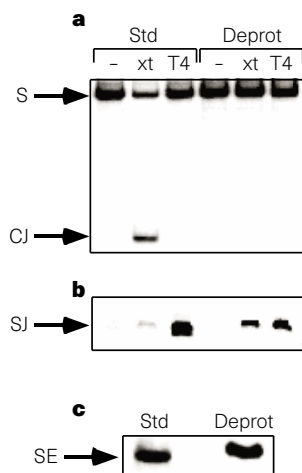


Figure 2 Effect of RAG proteins in joining reactions. In standard (std) reactions, RAG proteins were present throughout; in deproteinized (deprot.) reactions, RAG proteins were removed after cleavage; the HeLa cell fraction was included (xt) or not (–), or 20 U T4 DNA ligase was added instead of the HeLa cell fraction (T4). **a**, Coding junctions; **b**, signal junctions; **c**, SE (double-cut signal end), a DNA fragment produced by cleavage at both 12 and 23 type signals. Samples are SE DNA from (–) reactions (stained directly; see Methods).

deletion (marked “c” in Fig. 3a, b) and a minor product of 307 bp, representing junctions with 12 bp of P nucleotide addition (marked “a”). Joining of both these products is directed by 5-bp homologies (see above). Addition of recombinant human DNA ligase I to the standard reaction changed the size of the preferred ligation product. The major species now migrated between the two homology-directed joining products (marked “b”, lane 8 in Fig. 3a, and lanes 3 and 4 in Fig. 3b) at a position consistent with joints that are no longer homology directed. This same species could also be generated by addition of ligase I which was partially purified from HeLa cells but not frozen during preparation (data not shown).

Sequencing confirmed that the structure of the junctions made in the presence ligase I was altered as expected. Only 20% of these sequenced junctions used the 5-bp homologies (Table 1b). Coding junction sequences now included several examples of joints without deletion, as well as junctions with short deletions and P additions, in a pattern highly consistent with coding junctions derived from cells. The effects of DNA ligase I addition were not observed in reactions supplemented instead with T4 DNA ligase (Fig. 3a, lanes 3 and 4), recombinant human DNA ligase III (Fig. 3a, lanes 5 and 6), or DNA ligase IV purified from HeLa cells (Fig. 3b, lanes 5 and 6). The specificity of ligase I in altering coding joint structure suggests that this enzyme may be important for joining of *V(D)J* recombination intermediates.

It has previously been reported that a cell line having only mutant DNA ligase I still supports *V(D)J* recombination^{17,18}. Residual DNA ligase I activity (estimated to be 5–10% of normal levels) in this cell line may be sufficient for normal *V(D)J* recombination, as this level is sufficient to support DNA replication¹⁹. Alternatively, homology-independent joining in the absence of DNA ligase I may use an accessory factor or an as-yet unidentified mammalian DNA ligase.

Another important source of diversity in *V(D)J* rearrangements is the addition of non-templated nucleotides (N addition), which is mediated by terminal deoxynucleotidyl transferase (TdT) (reviewed in ref. 2). When we added TdT and ligase I to the cell-free reaction, most coding junctions contained N additions (Table 1c). We found a much wider range of N additions (2–54 bp) than in

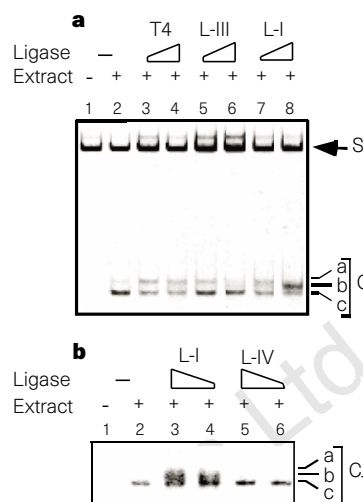


Figure 3 a, b, Effect of DNA ligases on formation of coding junctions. T4, T4 DNA ligase. L-III, recombinant human DNA ligase III. L-IV, human DNA ligase IV purified from HeLa cells. L-I, recombinant human DNA ligase I. CJ-a, 307-bp coding junction product. CJ-b, approximately 295-bp coding junction product. CJ-c, 283-bp coding junction product. CJ-a was visible only after long exposures (**b**). Each ligase was tested at two concentrations differing by a factor of 5. The activities of human DNA ligases added were similar, as determined by an oligonucleotide nick-sealing assay, whereas the activities of T4 DNA ligase (50 and 250 U) were at least 4-fold higher.

junctions derived from cells (typically 1–15 bp)².

Formation of signal junctions *in vivo* is rarely accompanied by base loss or addition, irrespective of the presence of TdT². To determine whether this is also true under cell-free conditions, we made use of the observation that precise formation of signal junctions generates a site for the restriction enzyme *Apa*LI. When PCR products of signal junctions were tested by *Apa*LI digestion, less than half of the signal junctions were precise (sensitive to *Apa*LI) in the absence of added ligase I, but this frequency was not affected by addition of TdT (Fig. 4b). Addition of ligase I increased both the efficiency (>10 fold; Fig. 4a) and accuracy of signal junction formation (Fig. 4b). Over 90% of junctions formed in the presence of ligase I were now precise, and again this frequency was unaffected by the presence of TdT (Fig. 4b). In contrast, the majority of the coding junctions from the same joining reaction had N additions (Table 1). We conclude that TdT activity is preferentially targeted to coding junctions in the cell-free reaction, as it is *in vivo*.

The absence of signal end modification by TdT may again be explained by their binding of RAG proteins. Otherwise, one would expect signal ends to be acceptable substrates for TdT addition. Blunt ends are known to be fully competent for TdT addition, comparable (within 2-fold) to substrates with 3' overhangs²⁰, as might be generated by the opening of hairpins.

The cell-free reaction described here reproduces the steps of *V(D)J* recombination quite accurately. When complete, the system carries out cleavage at two recombination signal sequence (RSS) sites, opens the resulting DNA hairpins, inserts or deletes nucleotides at these opened ends, and finally joins both coding and signal ends. The retention of P nucleotides at some coding junctions is also similar to the cellular process. We have shown that the RAG proteins not only perform cleavage, but have important effects on joining, providing evidence that the two stages of *V(D)J* recombination, cleavage and joining, are intimately linked. The fact that RAG proteins affect joining in a way specific to the type of end (signal or coding) has implications for *V(D)J* recombination. RAG proteins may confer directionality on the reaction (a signal end rarely joins to a coding end) and provide a basis for the different

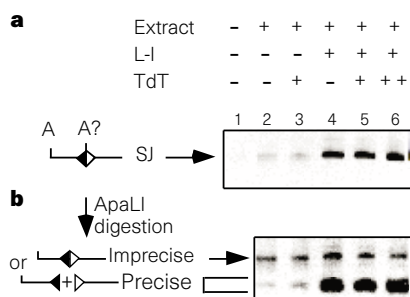


Figure 4 Effect of ligase I (L-I) and TdT on signal junction formation. 0.3 U (+) of TdT or 1.5 U (++) of TdT was added to reactions as indicated. **a**, 0.5% of reactions without, and 0.1% of reactions with DNA ligase I were analysed for signal junctions. A, *Apa*L1 site; A?, potential *Apa*L1 site. **b**, 60% of the PCR reactions in **a** were digested with *Apa*L1; the two digestion products of precise junctions co-migrate.

processing events at each type of end. Furthermore, persistence of a synaptic complex following cleavage could help to enforce joining of double-strand breaks originating from the same cleavage event, and reduce the potential for intermolecular (or interchromosomal) joining. Cell-free *V(D)J* recombination has identified a role for RAG proteins in the joining phase of the reaction and helped to distinguish the different effects of cellular DNA ligases. It may now also be useful in determining the role of other factors in processing hairpins and in the repair of double-strand breaks, including Ku (XRCC5 and XRCC6), DNAPK_{CS} (*scid*, XRCC7) and XRCC4 (see ref. 21 for example). If some of these double-strand-break repair factors are essential for cell growth, a cell-free system may be the best means for characterizing them. □

Methods

Reaction conditions. 50 ng plasmid pDVG54, which consists of the *Bam*HI/*Sal*I fragment of pDVG42 with recombination signals²² cloned into pUC18, was cleaved by ~200 ng RAG1 and RAG2 proteins⁶ in the presence of 40 ng HMG1 protein. Cleavage reactions were done in 10 µl at 37 °C for 1 h in 25 mM MOPS, pH 7.0, 2 mM DTT, 30 mM potassium glutamate, 30 mM KCl, 1% glycerol, 100 µg ml⁻¹ BSA, 4 mM MgCl₂ and 0.4 mM MnCl₂. About 70% of the DNA was cleaved under these conditions. In 'standard' reactions, joining factors were directly added to cleavage reactions; 'deproteinized' reactions were treated after the cleavage step with proteinase K (Gibco-BRL), extracted with phenol and chloroform, precipitated with ethanol, and reconstituted in cleavage buffer with HMG1 but without RAG proteins. Joining reactions were done overnight (16–20 h) at 25 °C with 500 ng HeLa cell extract, 0.5 mM ATP (Pharmacia) and 50 µM of each dNTP (Pharmacia) in 20 µl containing 10% w/v polyethylene glycol (average relative molecular mass 8K), and supplemented as necessary to maintain 25 mM MOPS, pH 7.0, 4 mM MgCl₂, 30 mM potassium glutamate. Reactions were stopped by addition of Tween-20 (Sigma) to 0.5%. Equivalent cleavage and recovery of substrate DNA in standard and deproteinized reactions was confirmed by electrophoresis and direct staining of DNA with SYBR green dye (FMC Bioproducts) from reactions without added extract or T4 ligase.

Analysis of joining reactions. Coding junctions were detected by PCR amplification of 20% of each joining reaction for 25 cycles using the primers DAR147 (5'-CATTCGCCATTCAGGCTGCGCAACT) and DAR5 (5'-TGCTTCCGGCTCGTATGTTGGTGAAT). Each cycle consisted of 15 s at 94 °C, 30 s at 65 °C, and 15 s at 72 °C. Signal junctions were detected by amplification of 0.1% (Fig. 4, lanes 4–6 only) or 0.5% (Fig. 2b and Fig. 4, lanes 1–3) of each joining reaction as described above, but with the primers DG88 (5'-GTAGAACGAGCGGCGT CGAAGCCTG) and DG89 (5'-CACAGCAGCGCCATATCGAAGGTCG). One third of each PCR reaction was analysed by non-denaturing polyacrylamide gel electrophoresis and visualized either by staining with ethidium bromide (Fig. 3a) or by Southern transfer and hybridization with an oligonucleotide probe for the 23 signal coding flank

(DAR153: 5'-CCCCGGTACCGAGCTC). Faint bands of the wrong size were not detected by Southern analysis and so were considered to be PCR artefacts. Serial dilutions of template over a 10-fold range produced proportional amounts of PCR product after amplification, indicating that the PCR reaction is in a linear range. PCR products were quantified in this range in at least two independent experiments, using ImageQuaNT software. For sequencing, coding joint PCR products were separated from amplified non-recombinant substrate by agarose gel electrophoresis, diluted 200-fold, and re-amplified as before, but for only 12–15 cycles and substituting the primer DAR149 (5'-CGCCAGCTGG CGAAAGGGGGATGT) for DAR147. Products from the secondary amplification were cloned using the TA cloning kit (Invitrogen) and sequenced using a plasmid Quick Denature Sequenase kit (Amersham).

HeLa extract. A cell extract was prepared by homogenization of 9.0×10^9 HeLa cells in buffer A (0.5 M KCl, 50 mM Tris, pH 8.0, 1 mM DTT, 10% glycerol, 1 mM PMSF and 1 µg ml⁻¹ each of leupeptin and aprotinin), followed by centrifugation at 100,000g. The supernatant was made 1 M in KCl, and nucleic acids were precipitated with 0.05% polyethyleneimine (PEI). To remove excess PEI, the supernatant was incubated with phosphocellulose, which was then removed by filtration. Material recovered in a 20–40% ammonium sulphate cut was resuspended in buffer B (0.1 M KCl, 20 mM Tris, pH 8.0, 1 mM DTT, 10% glycerol, 1 mM PMSF), dialysed, and loaded onto a ~15 ml Q-Sepharose column (Pharmacia). Bound protein was eluted with a gradient of 0.1–0.5 M KCl; active fractions eluted between 0.25 and 0.35 M KCl and were stored at -70 °C.

DNA ligases. Clones of histidine-tagged versions of human DNA ligases I and III (gift from T. Lindahl) were expressed in *E. coli* and purified by metal affinity. DNA ligase IV, purified from HeLa cells and free of the other human DNA ligases¹⁴, was also a gift from T. Lindahl. The activity of T4 DNA ligase (New England Biolabs) is defined in cohesive end units. The relative activity of each ligase was assessed using an oligonucleotide nick-joining assay.

Received 30 December 1996; accepted 22 May 1997.

- Gellert, M. Molecular analysis of *V(D)J* recombination. *Annu. Rev. Genet.* **22**, 425–446 (1992).
- Lewis, S. M. The mechanism of *V(D)J* joining: Lessons from molecular, immunological, and comparative analyses. *Adv. Immunol.* **56**, 27–150 (1994).
- Roth, D. B., Nakajima, P. B., Menetski, J. P., Bosma, M. J. & Gellert, M. *V(D)J* recombination in mouse thymocytes: double-strand breaks near T cell receptor δ rearrangement signals. *Cell* **69**, 41–53 (1992).
- Roth, D. B., Menetski, J. P., Nakajima, P. B., Bosma, M. J. & Gellert, M. *V(D)J* recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in *scid* mouse thymocytes. *Cell* **70**, 983–991 (1992).
- van Gent, D. C. *et al.* Initiation of *V(D)J* recombination in a cell-free system. *Cell* **81**, 925–934 (1995).
- McBlane, J. F. *et al.* Cleavage at a *V(D)J* recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83**, 387–395 (1995).
- Cortes, P. *et al.* *In vitro V(D)J* recombination: Signal joint formation. *Proc. Natl. Acad. Sci. USA* **93**, 14008–14013 (1996).
- Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, adjacent genes that synergistically activate *V(D)J* recombination. *Science* **248**, 1517–1523 (1990).
- Ramsden, D. A. & Gellert, M. Formation and resolution of double strand break intermediates in *V(D)J* rearrangement. *Genes Dev.* **9**, 2409–2420 (1995).
- Agrawal, A. & Schatz, D. G. RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in *V(D)J* recombination. *Cell* **89**, 43–53 (1997).
- Arrand, J. E., Willis, A. E., Goldsmith, I. & Lindahl, T. Different substrate specificities of the two DNA ligases of mammalian cells. *J. Biol. Chem.* **261**, 9079–9082 (1986).
- Elder, R. H., Montecucco, A., Ciarrocchi, G. & Rossignol, J.-M. Rat liver DNA ligases. *Eur. J. Biochem.* **203**, 53–58 (1992).
- Robbins, P. & Lindahl, T. DNA ligase IV from HeLa cell nuclei. *J. Biol. Chem.* **271**, 24257–24261 (1996).
- Wei, Y.-F. *et al.* Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol. Cell. Biol.* **15**, 3206–3216 (1995).
- Wang, Y.-C. J. *et al.* Mammalian DNA ligase II is highly homologous with vaccinia DNA ligase. *J. Biol. Chem.* **269**, 31923–31928 (1994).
- Tomkinson, A. E., Lasko, D. D., Daly, G. & Lindahl, T. Mammalian DNA ligases: catalytic domain and size of DNA ligase I. *J. Biol. Chem.* **265**, 12611–12617 (1990).
- Hsieh, C.-L., Arlett, C. F. & Lieber, M. R. *V(D)J* recombination in ataxia telangiectasia, Bloom's syndrome, and a DNA ligase I-associated immunodeficiency disorder. *J. Biol. Chem.* **268**, 20105–20109 (1993).
- Petrini, J. H. H., Donovan, J. W., Dimare, C. & Weaver, D. T. Normal *V(D)J* coding junction formation in DNA ligase I deficiency syndromes. *J. Immunol.* **152**, 176–183 (1994).
- Petrini, J. H. J., Xiao, Y. & Weaver, D. T. DNA ligase I mediates essential functions in mammalian cells. *Mol. Cell. Biol.* **15**, 4303–4308 (1995).
- Robbins, D. J. & Coleman, M. S. Initiator role of double stranded DNA in terminal transferase catalyzed polymerization reactions. *Nucleic Acids Res.* **16**, 2943–2957 (1988).
- Weaver, D. T. What to do at an end: DNA double strand break repair. *Trends Genet.* **11**, 388–392 (1995).
- van Gent, D. C., Ramsden, D. A. & Gellert, M. The RAG1 and RAG2 proteins establish the 12/23 rule in *V(D)J* recombination. *Cell* **85**, 107–113 (1996).

Acknowledgements. We thank T. Lindahl for purified human ligase IV and for clones expressing recombinant human DNA ligases I and III, and our colleagues for helpful discussions. D.A.R. was supported by fellowships from the MRC of Canada and the Cancer Research Institute, USA, and T.T.P. was supported by a Helen Hay Whitney postdoctoral fellowship.

Correspondence and requests for materials should be addressed to M.G. (e-mail: gellert@helix.nih.gov).

Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells

Ulf Grawunder^{*†}, Matthias Wilm^{*‡}, Xiantuo Wu^{*†}, Peter Kulesza^{*†}, Thomas E. Wilson^{*}, Matthias Mann[†] & Michael R. Lieber^{*†}

^{*} Washington University School of Medicine, Division of Molecular Oncology, Department of Pathology, 660 South Euclid Avenue, St Louis, Missouri 63110, USA

[‡] European Molecular Biology Laboratory (EMBL), Protein & Peptide Group, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

[†] Present address: Norris Comprehensive Cancer Center, University of Southern California, Rm 5425, Mailstop 73, 1441 Eastlake Avenue, Los Angeles, California 90033, USA

Mutation of the *XRCC4* gene in mammalian cells^{1,2} prevents the formation of the signal and coding joints in the *V(D)J* recombination reaction³, which is necessary for production of a functional immunoglobulin gene, and renders the cells highly sensitive to ionizing radiation⁴. However, *XRCC4* shares no sequence homology with other proteins, nor does it have a biochemical activity to indicate what its function might be². Here we show that DNA ligase IV (ref. 5) co-immunoprecipitates with *XRCC4* and that these two proteins specifically interact with one another in a yeast two-hybrid system. Ligation of DNA double-strand breaks in a cell-free system by DNA ligase IV is increased fivefold by purified *XRCC4* and seven- to eightfold when *XRCC4* is co-expressed with DNA ligase IV. We conclude that the biological consequences of mutating *XRCC4* are primarily due to the loss of its stimulatory effect on DNA ligase IV: the function of the *XRCC4*-DNA ligase IV complex may be to carry out the final steps of *V(D)J* recombination and joining of DNA ends.

We immunoprecipitated *XRCC4* from mammalian cells in order to identify factors that might be associated with it. We used the *XRCC4*-deficient Chinese hamster ovary cell line XR-1 (ref. 1), which has been stably transfected with an expression vector encoding human *XRCC4* fused to a 9× histidine tag and a 3× Myc epitope at the carboxy terminus. The tagged *XRCC4* protein was detectable by anti-Myc western blotting as a double band, with the apparent relative molecular mass (M_r) of each band being 60K and 65K (Fig. 1). The existence of a double band of this size for *XRCC4* suggests that there must be extensive post-translational modification, because the calculated M_r of the tagged *XRCC4* protein is 44K. Four different stable *XRCC4* transfectants showed full reconstitution to wild-type *V(D)J* recombination (not shown) and resistance to X-rays (Fig. 1), indicating that human *XRCC4* complementary DNA functionally compensates for the deleted hamster *XRCC4* gene.

Immunoprecipitation of each of the four stable *XRCC4* transfectants gave a protein band of $M_r \sim 95K$ that co-immunoprecipitated with epitope-tagged *XRCC4* protein (Fig. 2: XAP is *XRCC4*-associated protein). This XAP band was not co-immunoprecipitated using XR-1 cell lysates or other His/Myc tagged proteins (such as Rag-1 and -2; data not shown). The *XRCC4*/XAP protein complex remained stable up to a salt concentration of 1.2 M KCl, indicating that this protein-protein interaction was very stable.

XAP was isolated from silver-stained gels and microsequenced as described⁶. The sequences of three tryptic peptides for XAP were determined, which unequivocally identified XAP as DNA ligase IV: one octapeptide originated from the amino terminus (amino acids

2–9), one octapeptide had an identical sequence to that between residues 447–454, and one nonapeptide was from the C-terminal tail (amino acids 771–779). Two of these three regions are not well conserved among ligases I, III and IV (ref. 5), and yet all 25 amino-acid residues were identical to the sequence of human DNA ligase IV, suggesting a high degree of homology between rodent and human DNA ligase IV. Single-strand-nick ligase activity was reproducibly detectable in immunoprecipitates from *XRCC4*-transfected cells but was absent in immunoprecipitates from XR-1 control cells (data not shown). This activity was independent of the addition of exogenous ATP, which is a feature of human DNA ligase IV (ref. 7).

We next tested whether the interaction between *XRCC4* and DNA ligase IV could be reproduced in the yeast two-hybrid system. For this and all further experiments, human DNA ligase IV cDNA was cloned from Reh cDNA using the polymerase chain reaction (PCR). DNA ligase IV was found specifically to interact with *XRCC4*, but not with two selected control proteins (the 70K and 86K subunits of the DNA-end-binding Ku heterodimer⁸; Table 1). This finding independently demonstrates that DNA ligase IV interacts with *XRCC4* and that this occurs in a eukaryotic cell nucleus.

All known eukaryotic DNA ligases form covalent AMP-enzyme adducts upon addition of ATP⁹. However, in contrast to other DNA ligases DNA ligase IV is pre-adenylated in mammalian cells⁷. To determine whether *XRCC4* participates in this pre-adenylation, human DNA ligase IV was expressed as a His/Myc-tagged protein in XR-1 cells using a vaccinia virus overexpression system. Its adenylation *in vitro* was compared with that of endogenous DNA

Table 1 Interaction of *XRCC4* with DNA-ligase IV in yeast two-hybrid system

LexA fusion	B42 fusion	Growth on Leu ⁻		β -Galactosidase activity		Activity ratio
		Glu	Gal	Glu	Gal	
hDNL4	XRCC4	–	+	2.2	54.4	24.4
48K Ku70	XRCC4	–	–	1.7	6.3	3.7
Ku86	XRCC4	–	–	2.1	4.5	2.1
Bicoid	XRCC4	–	–	2.3	7.8	3.4
hDNL4	Ku70	–	–	2.4	5.7	2.4
hDNL4	Ku86	–	–	2.0	6.2	3.1

XRCC4 expressed as a B42-transactivation fusion protein was tested for its ability to interact with various LexA fusion proteins. Ku70 and Ku86 fusion constructs were used as controls. A B42 transactivation construct not carrying a fusion partner was included in each set as an additional control. Gal, galactose; Glu, glucose.

* β -Galactosidase activity is expressed in units per 10^7 cells.

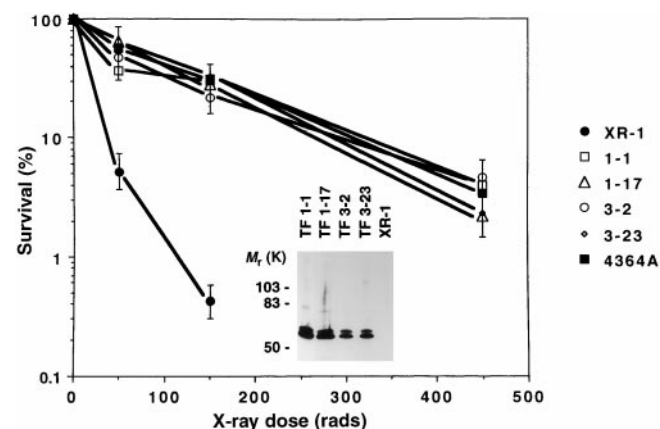


Figure 1 Functional characterization of stable *XRCC4* transfectants. Stable *XRCC4*-His₉/Myc₃ transfectants were generated in the CHO cell line XR-1, which is deficient for *XRCC4* (ref. 2). X-ray sensitivity for four different stable *XRCC4* transfectants (designated (TF) 1-1, 1-17, 3-2, 3-23) as well as the wild-type parental cell line of XR-1, 4364A, are shown. Expression of the stably transfected *XRCC4* protein assayed by anti-Myc western blotting is shown in the insert.

ligase IV in complex with XRCC4. Both the recombinant and endogenous ligases could be efficiently adenylated only after pre-treatment with inorganic pyrophosphate (PP_i), which strips the AMP–ligase adduct (Fig. 3). This indicates that recombinant DNA ligase IV is pre-adenylated as extensively as endogenous DNA ligase IV and that XRCC4 is therefore not needed for pre-adenylation of DNA ligase IV.

Purified human DNA ligase IV was previously believed to be able to seal only single-strand DNA breaks and not double-strand breaks⁷. Using an oligonucleotide-based intramolecular ligation assay (Fig. 4a), we tested the activity of the XRCC4–DNA ligase IV complex under different conditions. Nick-ligation activity (on single-strand breaks) was quite insensitive to variations in salt concentration, divalent cations or temperature. In contrast, duplex ligations required specific conditions that were optimal in 100 mM KCl at 37 °C with Mn²⁺ present (data not shown). Under these conditions, linearized plasmids with compatible ends could also be religated efficiently (data not shown). Duplex ligation by XRCC4–DNA ligase IV was detectable at blunt DNA ends, but was much more efficient at sticky ends (Fig. 4a). As the repair of DNA double-strand breaks frequently involves short regions of micro-homology (as in V(D)J recombination¹⁰), our duplex ligation assay with compatible ends should mimic non-homologous DNA-end-joining events *in vivo*.

Using this system, we compared DNA ligase IV activity on sticky ends in the presence and absence of XRCC4 (Fig. 4b). Co-expression with XRCC4 reproducibly increased the activity of DNA ligase IV 7–8-fold, as determined by phosphorimager analysis (Fig. 4b). This effect is partly due to a consistent slight increase in the steady-state level of DNA ligase IV (~2-fold; data not shown) that carries over in immunoprecipitation (Fig. 4c), but the difference in activity always exceeded the observed differences in DNA ligase IV levels. We therefore tested whether adding purified recombinant XRCC4

could stimulate DNA ligase IV and found that recombinant XRCC4 at a molar ratio of 1:1 consistently gave a 5-fold increase in sticky-end ligation (Fig. 4b). XRCC4 protein itself had no ligase activity, and a nonspecific protein (bovine serum albumin) had no effect on DNA ligase IV.

Our results clarify the role of XRCC4 in DNA metabolism. They indicate XRCC4 is needed to stimulate DNA ligase activity, primarily through direct structure interaction with DNA ligase IV. This interaction distinguishes XRCC4 from the DNA-end-binding auto-antigen Ku70/Ku86 (refs 11, 12) and from DNA-PK (ref. 13), a kinase that is mutated in severe combined immunodeficiency^{14,15}, because these activities appear to be required earlier on for recognition (Ku70/86) and processing of DNA ends (DNA-PK)¹⁶.

A logical extension of this is that DNA ligase IV, in complex with XRCC4, is the joining activity in V(D)J recombination and generally in DNA-end joining. Previous work with mammalian cell lines with deficiencies in DNA ligase I and III indicates that they are entirely normal for V(D)J recombination^{3,17–19}. The DNA ligase II cDNA has not been identified, but it may represent an alternatively spliced form of DNA ligase III or a start-codon variant^{5,20}. Thus, DNA ligases I, II and III are unlikely to be the ligases involved in V(D)J recombination and DNA-end joining. Our conclusion that DNA ligase IV is the end-joining activity is supported by results in an accompanying paper²¹, in which the deletion of the homologue of DNA ligase IV in yeast severely impairs non-homologous DNA-end joining. It is likely that a deficiency in either XRCC4 or DNA ligase IV results not only in an increased sensitivity of cells to X-rays, but may also give rise to immunodeficiency in animals. □

Methods

Expression vectors. Human XRCC4 cDNA (gift from F. Alt) was cloned into the mammalian expression vector pCDNA3 and fused to a sequence encoding a C-terminal His₆/Myc₃ tag (plasmid pUG14). HA-tagged XRCC4 was generated

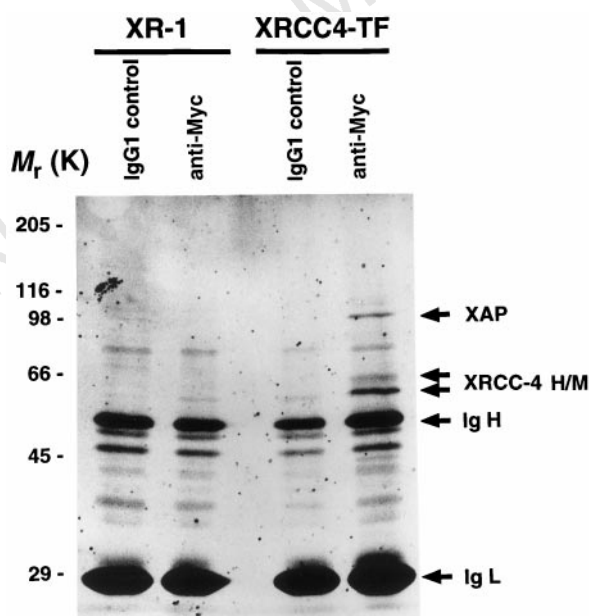


Figure 2 Immunoprecipitation (IP) of epitope-tagged XRCC4. Preparative anti-Myc IPs from total cell lysates (10⁸ cells per lane) using XR-1 and XRCC4 transfectant cell line 3-2. Immunoprecipitated proteins were visualized using SYPRO orange stain. IPs from XRCC4 transfectant cells show the 60K/65K XRCC4 protein bands and co-immunoprecipitate of one additional band, XAP, migrating at ~95K. IP controls were carried out with an IgG₁ control antibody (mAb MR5-2, specific for mouse TCR Vβ8). Bands migrating at 50K and 25K–30K are the heavy and light chains of the Myc mAb, respectively.

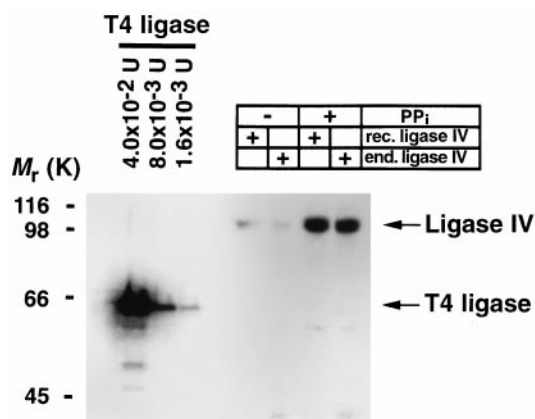


Figure 3 Adenylation of recombinant and endogenous DNA ligase IV. Recombinant, 9 × His/3 × Myc-tagged human DNA ligase IV was transiently expressed in XR-1 cells using the vaccinia virus expression system. Endogenous DNA ligase IV was isolated by co-immunoprecipitation from stable XRCC4 transfectants. Both the endogenous and recombinant DNA ligase IVs were adenylated with [α-³²P]ATP, either with or without pyrophosphate pre-treatment (±PP_i, as indicated). *In vitro* adenylated commercial T4 DNA ligase was used as a positive control.

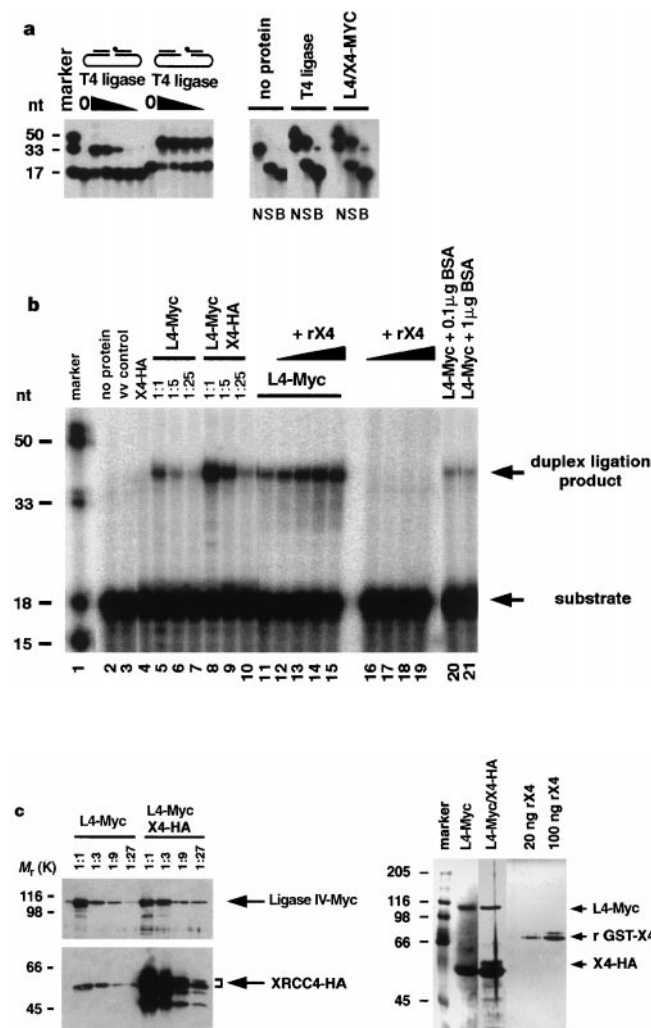


Figure 4 XRCC4 stimulates DNA ligase IV activity *in vitro*. **a**, DNA oligonucleotide substrates for blunt-end and compatible-end ligations were created by annealing a DNA 80-mer to two short oligomers that were complementary to each end of the long oligomer. One of the short oligomers was 5'-labelled (black dot), resulting in incorporation of the label upon ligation of the two double-stranded arms. Left, analysis by denaturing PAGE of serial 10-fold dilutions of T4 DNA ligase starting at 1U per reaction. Right, ligation with nicked (N), sticky-end (S) or blunt-end (B) substrates catalysed by either 5×10^{-3} U T4 DNA ligase or DNA ligase IV/XRCC4 co-immunoprecipitated from 10^7 XRCC4 transfectants. Lanes with no protein indicate the position of each substrate. **b**, Sticky-end ligation catalysed by DNA ligase IV alone (L4-MYC) or by DNA ligase IV in complex with XRCC4 (L4-MYC/X4-HA, and L4-MYC + rX4), analysed by denaturing PAGE. In one experiment, DNA ligase IV was expressed in XR-1 cells either in the absence (lanes 5–7) or presence (lanes 8–10) of the XRCC4 expression vector. Proteins were isolated by anti-Myc IP. Ligations with serial 5-fold dilutions are shown. In a second experiment, DNA ligase IV expressed in the absence of XRCC4 (lane 11) was supplemented with increasing amounts of recombinant XRCC4 protein (lanes 12–15), ending at a molar ratio at 1:1 (lanes 12–15). Control ligations were carried out with anti-Myc IPs from vaccinia-virus-infected (lane 2) or XRCC4-HA-expressing XR-1 cells (lane 3), as well as with identical amounts of recombinant XRCC4 protein alone (lanes 16–19). A possible nonspecific protein effect on DNA ligase IV activity was assayed by addition of BSA (lanes 20, 21). **c**, Anti-Myc (top/left) and anti-HA western blot (bottom/left) analysis of serial dilutions of immunoprecipitated proteins used in ligation assays, and SYPRO orange staining of the proteins (right). GST-tagged recombinant XRCC4 from insect cells is designated as GST-X4. Note that HA-tagged XRCC4 migrates close to the heavy chain of the antibody used for IP.

by replacing the Myc epitope in pUG14 with a linker encoding the HA epitope (plasmid pUG23). Human DNA ligase IV was PCR-cloned from Reh cDNA into pCDNA3 also containing a C-terminal His/Myc tag (plasmid pUG19). The XRCC4-His₉/Myc₃ gene was cloned into baculovirus vector pAcG2T (Pharmingen) from pUG14 in-frame with an N-terminal glutathione-S-transferase (GST) tag (pUG10).

Stable XRCC4 transfectants and XRCC4 baculovirus. Stable puromycin-resistant colonies were generated by co-transfecting plasmid pUG14 with a puromycin-resistance-marker-containing plasmid LXSP. Transfectants were selected at $4 \mu\text{g ml}^{-1}$ puromycin. Recombinant XRCC4 baculovirus was generated by co-transfecting pUG10 with Baculo Gold transfection module (Pharmingen), as recommended by the manufacturer.

Immunoprecipitation. Cells were lysed in immunoprecipitation (IP) buffer (25 mM HEPES/KOH, 0.25% Tween-20, 10 mM MgCl₂, 300 mM KCl, pH 7.9, supplemented with 1 mM PMSF, $2 \mu\text{g ml}^{-1}$ aprotinin, $2 \mu\text{g}^{-1}$ leupeptin and $1 \mu\text{g ml}^{-1}$ pepstatin A). Immunoprecipitates were prepared with mAb MYC 1-9E10.2 (ATCC CRL1729) and protein G-Sepharose and were washed 5 times with 25 mM HEPES/KOH, 0.1% NP40, 10 mM MgCl₂, 650 mM KCl, pH 7.9.

Proteins were fractionated by SDS-PAGE, analysed with SYPRO orange stain and analysed on a STORM phosphorimager (Molecular Dynamics). ECL western blots have been described²². Anti-Myc and anti-HA blots were developed with mAb 1-9E10.2 and mAb 3F10 (Boehringer Mannheim), respectively, followed by horseradish-peroxidase-conjugated secondary reagents.

X-ray sensitivity and V(D)J recombination assays. Cells grown to confluency were γ -irradiated and plated into 96-well plates under limiting dilution conditions. The percentage surviving was calculated by dividing the number of colonies at a given dilution of irradiated cells by the number of colonies at the

same dilution of non-irradiated cells. Transient cellular V(D)J recombination was assayed as described^{23,24}. Recombination substrates retaining either coding or signal joints upon V(D)J recombination were pGG51 and pGG49, respectively²⁵.

Protein microsequencing. Microsequencing of tryptic peptides from silver-stained protein bands (500–1,000 ng XAP) was done by nano-electrospray tandem mass spectrometry⁶. From peptide sequences, DNA ligase IV was uniquely identified in a non-redundant database containing more than 200,000 protein sequences without using species or protein size restrictions²⁶.

Ligation assay and *in vitro* adenylation. Nick ligation and adenylation were done in 60 mM Tris 10 mM MgCl₂, 5 mM DTT, $5 \mu\text{g ml}^{-1}$ acetylated BSA, pH 7.9. Ligations were done with 0.1 pmol nick-ligation substrate (a 50-nucleotide bottom strand annealed to a complementary 17-mer and a 5'-phosphorylated 33-mer forming the nicked top strand). Ligations were incubated for 30 min at room temperature, denatured at 95 °C in formamide, and fractionated on denaturing 20% polyacrylamide gels. Adenylation was done with $10 \mu\text{Ci}$ [α -³²P]ATP for 15 min at room temperature. Pyrophosphate pretreatment was for 15 min at RT with 5 mM Na₂P₂O₇.

Duplex ligations were carried out in 50 mM HEPES, 100 mM KCl, 5 mM MnCl₂, 2 mM 2-mercaptoethanol, pH 7.9. Primer sequences for duplex ligation substrates: 80-mer UG301: 5'-GGGTGGATTCGTCGCTTTCCTTCCCTTTTCTCTCTTTTCCCTTTCCTTCCTTCCTTCCCTCCCTCGTCTAGATCCC-3', two 15-mers UG302: 5'-CGGACGAATCCACCC-3' and 5'-[³²P]UG303: 5'-GGGATCTAGACGAGG-3' (for blunt-end substrate) and two 18-mers UG304: 5'-CGGACGAATCCACCCGGG-3' and 5'-[³²P]UG305: 5'-ATCTAGACGAGGGGAGGG-3' (for sticky-end substrate). Duplex ligations were incubated for 1 h at 37 °C.

Expression and purification of recombinant proteins. Recombinant GST-tagged XRCC4 was purified from Hi-5 insect cells infected with baculovirus bVUG7-3. Cells were lysed in IP buffer containing 20 mM imidazole. Proteins were bound to Ni-NTA agarose (Qiagen) and eluted with buffer containing 100 mM EDTA. Supernatants were then incubated with glutathione–Sephadex 4B (Pharmacia) which was eluted with buffer containing 10 mM reduced glutathione. Proteins were dialysed overnight at 4 °C against 50 mM HEPES, 100 mM KCl, pH 7.9.

Recombinant Myc-tagged DNA ligase IV was expressed for 24 h in XR-1 cells by liposome-mediated transfection of the appropriate expression vector (10 µg plasmid per 3×10^7 cells) and co-infection with vaccinia virus vTF7-3 (ref. 27).

Yeast-two-hybrid system. Protein interactions were analysed using the yeast-two-hybrid system as described^{28–30}. The Ku70 and Ku86 fusion constructs have been reported⁸.

Received 19 December 1996; accepted 14 May 1997.

1. Stamato, T. D., Weinstein, R., Giaccia, A. & Mackenzie, L. Isolation of cell-cycle dependent gamma-ray sensitive Chinese hamster ovary cell. *Somat. Cell Mol. Genet.* **9**, 165–173 (1983).
2. Li, Z. *et al.* The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83**, 1079–1089 (1996).
3. Taccioli, G. E. *et al.* Impairment of V(D)J recombination in double-strand break repair mutants. *Science* **260**, 207–210 (1993).
4. Giaccia, A. J., Richardson, E., Denko, N. & Stamato, T. D. Genetic analysis of the XR-1 mutation in hamster and human hybrids. *Somatic Cell Mol. Genet.* **15**, 71–79 (1989).
5. Wei, Y.-F. *et al.* Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol. Cell. Biol.* **15**, 3206–3216 (1995).
6. Wilm, M. *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469 (1996).
7. Robins, P. & Lindahl, T. DNA ligase IV from HeLa cell nuclei. *J. Biol. Chem.* **271**, 24257–24261 (1996).
8. Wu, X. & Lieber, M. R. Protein–protein and protein–DNA interaction regions within the DNA end binding protein Ku70–Ku86. *Mol. Cell. Biol.* **16**, 5186–5193 (1996).
9. Lindahl, T. & Barnes, D. E. Mammalian DNA ligases. *Annu. Rev. Biochem.* **61**, 251–281 (1992).
10. Gerstein, R. M. & Lieber, M. R. Extent to which homology can constrain coding exon junctional diversity in V(D)J recombination. *Nature* **363**, 625–627 (1993).
11. Reeves, W. H. & Stoeber, Z. M. Molecular cloning of cDNA encoding the p70(Ku) lupus autoantigen. *J. Biol. Chem.* **264**, 5047–5052 (1989).
12. Mimori, T. *et al.* Isolation and characterization of cDNA encoding the 80-kDa subunit protein of the human autoantigen Ku (p70/80) recognized by autoantibodies from patients with scleroderma–polymyositis overlap syndrome. *Proc. Natl Acad. Sci. USA* **87**, 1777–1781 (1990).
13. Hartley, K. O. *et al.* DNA-dependent protein kinase catalytic subunit: A relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* **82**, 849–856 (1995).
14. Kirchgessner, C. U. *et al.* DNA-dependent kinase (p350) as a candidate gene for the murine SCID defect. *Science* **267**, 1178–1183 (1995).
15. Blunt, T. *et al.* Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell* **80**, 813–823 (1995).
16. Jackson, S. P. & Jeggo, P. A. DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK. *Trends Biochem. Sci.* **20**, 412–415 (1995).
17. Hsieh, C. L., Arlett, C. F. & Lieber, M. R. V(D)J recombination in ataxia telangiectasia, Bloom's syndrome, and a DNA ligase I-associated immunodeficiency disorder. *J. Biol. Chem.* **268**, 10105–10109 (1993).
18. Petrini, J., Donovan, J. W., Dimare, C. & Weaver, D. T. Normal V(D)J coding junction formation in DNA ligase I deficiency syndromes. *J. Immunol.* **152**, 176–183 (1994).
19. Kemp, L., Sedgwick, S. & Jeggo, P. X-ray sensitive mutants of Chinese hamster ovary cells defective in double-strand rejoining. *Mut. Res.* **132**, 189–196 (1984).
20. Chen, J. *et al.* Mammalian DNA ligase III: molecular cloning, chromosomal localization, and expression in spermatocytes undergoing meiotic recombination. *Mol. Cell. Biol.* **15**, 5412–5422 (1995).
21. Wilson, T. E., Grawunder, U. & Lieber, M. R. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* **388**, 495–498 (1997).
22. Grawunder, U., Schatz, D. G., Leu, T. M., Rolink, A. & Melchers, F. The half-life of RAG-1 protein in precursor B cells is increased in the absence of RAG-2 expression. *J. Exp. Med.* **183**, 1731–1737 (1996).
23. Lieber, M. R. *et al.* The defect in murine severe combined immune deficiency: joining of signal sequences but not coding segments in V(D)J recombination. *Cell* **55**, 7–16 (1988).
24. Gauss, G. & Lieber, M. R. DEAE–dextran enhances electroporation of mammalian cells. *Nucleic Acids Res.* **20**, 6739–6740 (1992).
25. Miller, R. D. *et al.* Gene for the catalytic subunit of mouse DNA-PK maps to the scid locus. *Proc. Natl Acad. Sci. USA* **92**, 10792–10795 (1995).
26. Mann, M. & Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analyt. Biochem.* **66**, 4390–4399 (1994).
27. Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. Eukaryotic transient expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* **83**, 8122–8126 (1986).
28. Gyuris, J., Golemis, E., Chertkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791–803 (1993).
29. Zervos, A., Gyuris, J. & Brent, R. Mxi1, a protein that specifically interacts with Max to bind Myc–Max recognition sites. *Cell* **72**, 223–232 (1993).
30. Guarente, L. Yeast promoters and lacZ fusion designed to study expression of cloned genes in yeast. *Meth. Enzymol.* **101**, 181–191 (1983).

Acknowledgements. U.G. is a postdoctoral fellow of the Boehringer Ingelheim Foundation. T.E.W. is a Howard Hughes Medical Institute physician postdoctoral fellow. This work was supported by grants to M.R.L. M.R.L. is a Leukemia Society of America scholar. We thank C. Hsieh, A. Kalb and M. Yaneva for comments on the manuscript.

Correspondence and requests for materials should be addressed to M.R.L. (e-mail: lieber_m@froggy.hsc.usc.edu).

Yeast DNA ligase IV mediates non-homologous DNA end joining

Thomas E. Wilson[†], Ulf Grawunder^{†‡} & Michael R. Lieber^{†‡}

Washington University School of Medicine, * Division of Laboratory Medicine, † Division of Molecular Oncology, Department of Pathology, Box 8118, 660 S. Euclid Avenue, St Louis, Missouri 63110, USA

‡ Present address: Norris Comprehensive Cancer Center, University of Southern California, Rm 5425, Mailstop 73, 1441 Eastlake Avenue, Los Angeles, California 90033, USA

The discovery of homologues from the yeast *Saccharomyces cerevisiae* of the human Ku DNA-end-binding proteins (*HDF1* and *KU80*) has established that this organism is capable of non-homologous double-strand end joining (NHEJ)^{1–5}, a form of DNA double-strand break repair (DSBR) active in mammalian V(D)J recombination^{6–8}. Identification of the DNA ligase that mediates NHEJ in yeast will help elucidate the function of the four mammalian DNA ligases in DSBR, V(D)J recombination and other reactions^{9,10}. Here we show that *S. cerevisiae* has two typical DNA ligases, the known DNA ligase I homologue CDC9 (refs 11–14) and the previously unknown DNA ligase IV homologue DNL4. *dnl4* mutants are deficient in precise and end-processed NHEJ. *DNL4* and *HDF1* are epistatic in this regard, with the mutation of each having equivalent effects. *dnl4* mutants are complemented by overexpression of Dnl4 but not of Cdc9, and deficiency of Dnl4 alone does not impair either cell growth or the Cdc9-mediated responses to ionizing and ultraviolet radiation. Thus, *S. cerevisiae* has two distinct and separate ligation pathways.

We reasoned that yeast might have an unidentified DNA ligase that catalyses NHEJ, the direct rejoining of a double-strand DNA break in a non-homologous 'end-to-end' fashion, because mammalian cells deficient in DNA ligase I can perform V(D)J recombination^{15,16}. We therefore screened the completed *S. cerevisiae* genome sequence by computer analysis for the presence of DNA ligases using all known ligases as the query sequences. Two were revealed: CDC9 itself and the open reading frame YOR005c on the right arm of chromosome XV (accession number Z74913; $P < 7.6 \times 10^{-19}$ versus Cdc9). All other potential matches showed a considerably lower score ($P > 0.17$), with no clear presence of the invariant catalytic lysine. Thus, *S. cerevisiae* has two and only two typical ATP-dependent DNA ligases.

Multiple sequence alignment revealed that YOR005c is the yeast homologue of human DNA ligase IV (termed DNL4; Fig. 1), based on clear conservation of the catalytic domain, the reactive site lysine, and the unique carboxy-terminal extension of human DNA ligase IV (ref. 17). The second ATG in the overall human DNA ligase IV open reading frame has been assigned as the start codon¹⁷, but there is a meaningful alignment with Dnl4 starting with the first methionine, arguing that this larger open reading frame is used in some if not all mammalian cells.

All ATP-dependent DNA ligases form a covalent ligase–adenylate intermediate as part of the catalytic mechanism¹⁰. Only DNA ligase IV forms this complex intracellularly in a stable fashion, so that purified protein is adenylated completely with [α -³²P]ATP *in vitro* only after removal of bound AMP, for example by incubation with sodium pyrophosphate (pyrophosphorolysis)¹⁸. Likewise, covalent protein–adenylate complexes formed specifically with an immunoprecipitated glutathione S-transferase (GST)–Dnl4 fusion protein with almost 4-fold greater efficiency following preincubation with sodium pyrophosphate (Fig. 2), demonstrating that Dnl4 bears

functional as well as structural homology with human DNA ligase IV.

Yeast bearing the *dnl4Δ* mutation grew at rates comparable to wild type with no visibly apparent phenotypic change, indicating that Dnl4 is not required for the ligation of Okazaki fragments during DNA replication. Further, yeast bearing only this mutation did not show enhanced sensitivity to ultraviolet light (Fig. 3a) or ionizing radiation (Fig. 3b), indicating that Dnl4 is not required for either excision repair or RAD52-dependent homologous recombination, respectively. However, *rad52/dnl4* yeast did show an increased sensitivity to ionizing radiation compared with *rad52/DNL4* yeast, to an extent indistinguishable from that of *rad52/hdf1* yeast (that is, Ku-deficient; Fig. 3b). This is attributed to the loss of NHEJ in either *hdf1* or *dnl4* mutants which is revealed only when the dominant homologous DSB repair pathway is disabled by disruption of the *RAD52* gene^{3,5,19}. A *rad52/hdf1/dnl4* triple-mutant strain showed no further increase in radiosensitivity beyond the double mutants, demonstrating that Hdf1 and Dnl4 proteins operate in the same

DNA-repair pathway, namely NHEJ. Notably though, *dnl4* yeast cells were able to grow at 37°C, whereas *hdf1* yeast were not^{1-5,19}, indicating that Hdf1 is involved in at least one *DNL4*-independent cellular process.

We next modified an assay in which restriction enzyme-linearized plasmids are transformed into yeast and rejoined intracellularly^{2,3,5}, specifically by creating breaks in an *ADE2* gene carried on *URA3*-selectable plasmids. Plasmid pES26 has a polyterminator inserted into the *Bgl*III site of the *ADE2* gene. Yeast cells that have been transformed with *Bgl*III-digested pES26 and precisely rejoined the *ADE2* coding sequence (with no nucleotide loss or addition) can therefore be selected on adenine-lacking medium, because uncut plasmid and imprecise joins will yield *ade2* yeast. In contrast, plasmid pES16 lacks the polyterminator. Transforming *Bgl*III-digested pES16 therefore allows the simple visualization of colonies that have rejoined the plasmid imprecisely, because these appear red (that is, *ade2*) in a background of white (*ADE2*) colonies. With both substrates, the plasmid-borne *ADE2* gene can also be repaired by

Table 1 Homologous DSB repair versus precise NHEJ in yeast strains

Strain	GST-ligase	ADE2 locus genotype	
		<i>ade2-101</i>	<i>ade2Δ::TRP1</i>
Parent	–	100	36.2 ± 6.1
<i>hdf1</i>	–	12.3 ± 3.6	0.7 ± 0.7
<i>dnl4</i>	–	22.1 ± 8.8	0.7 ± 0.2
<i>dnl4</i>	Cdc9	ND	0.6 ± 0.5
<i>dnl4</i>	Dnl4	ND	33.0 ± 1.5
<i>rad52</i>	–	50.1 ± 16.4	30.1 ± 4.7

Yeast strains with the indicated combination of gene disruption and chromosomal *ADE2* genotype were transformed with *Bgl*III-digested pES26 and plated to minimal medium lacking uracil and adenine, selecting for precisely joined plasmids. Relative transformation efficiencies (%) are shown, with the reference strain for all entries being the *ade2-101* parent strain, YW104. All data are normalized to parallel transformations with supercoiled plasmid. Results are the mean ± standard deviation of two independent experiments. Where indicated, the *ade2Δ::TRP1/dnl4* strain was first transformed with a plasmid expressing a GST–ligase fusion protein from the strong constitutive *ADHI* promoter. ND, not determined.

Table 2 Imprecise NHEJ is impaired in *hdf1* and *dnl4* yeast

Strain	Relative efficiency of imprecise joining	<i>Pvu</i> II fragment sizes in kb	Fraction with macrodeletion
Parent	100	3.0 (12)	0/12
<i>hdf1</i>	36.2 ± 10.5	2.9, 2.6, 2.4 (3) 2.3 (2), 2.2 (4), 1.7	12/12 (<i>P</i> = 7.4 × 10 ^{−7})
<i>dnl4</i>	16.4 ± 4.4	2.8, 2.4, 2.2 (6) 1.7 (2), 1.6, 1.4	12/12 (<i>P</i> = 7.4 × 10 ^{−7})

The *ade2Δ::TRP1* parent strain YW112 and its derivatives with the indicated disruptions were transformed with *Bgl*III-digested pES16 and plated to minimal medium lacking uracil and containing 10 μg ml^{−1} adenine. Relative efficiencies (%) are calculated relative to YW112, counting only *ade2* (red) colonies as an indication of the rate of imprecise joining. Results are the mean ± standard deviation of 3 experiments. Several plasmids were then recovered from *ade2* transformants of each strain and the size of their *ADE2*-containing *Pvu*II fragments determined (numbers in parentheses indicate the number of times a given fragment size was observed if greater than one). For reference, the *Pvu*II fragment is 3.0 kb in pES16. *P* values (right column) are from Fisher's exact test.

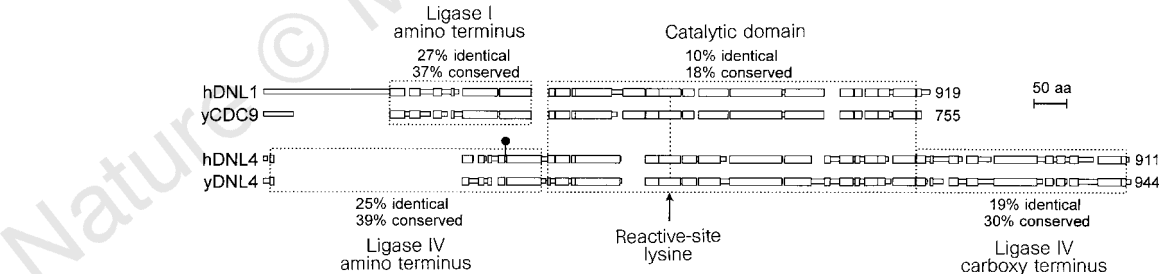


Figure 1 Multiple sequence alignment of human DNA ligase I (hDNL1), yeast CDC9, human DNA ligase IV (hDNL4) and yeast DNL4. Blocks of homology are represented by tall bars, unaligned (dissimilar) regions by short bars, and gaps by blank areas. The position of the conserved reactive-site lysine is indicated by an arrow. Boxes surround the large homologous domains, showing the per cent

amino-acid identity and per cent identity plus highly conservative substitution ('conserved') among all sequences within a given box/domain. The ball on a stick marks the position of the second human DNA ligase IV methionine, originally assigned as the start codon¹⁷.

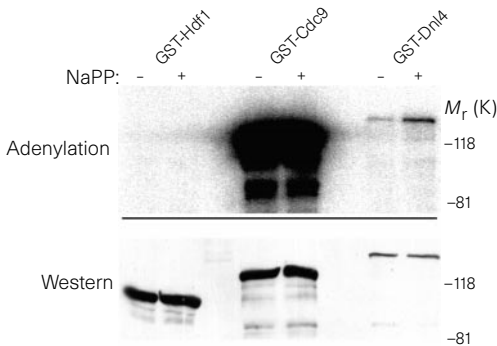


Figure 2 GST–Dnl4 is a pre-adenylated DNA ligase. GST fusion proteins were immunoprecipitated from whole-cell extracts and adenylated with [α -³²P]ATP. Where indicated, preincubation with sodium pyrophosphate (NaPP) was used to strip preformed ligase–adenylate complexes. Images were generated by Phosphorimager screen exposure (top) and western blotting (bottom) of the same nitrocellulose blot of adenylated proteins. Shorter exposure of the Phosphorimager screen revealed no difference in the extent of adenylation of the two GST–Cdc9 samples.

recombination with the chromosome, but we can prevent this homologous DSB event by deleting the chromosomal *ADE2* gene.

Table 1 shows the relative efficiency of precise NHEJ among yeasts with a variety of gene disruptions and chromosomal *ADE2* genotypes. All results are expressed relative to the parent strain after normalization for the slight variation in overall transformability as assessed by parallel transformations with supercoiled plasmid DNA. Disruption of either *HDF1*, *DNL4* or *RAD52* in a parent strain that has the *ade2-101* chromosomal mutation (an *ochre* nonsense mutation) partially reduced the efficiency of transformation, indicating that both homologous DSB and NHEJ pathways are active in the parent strain. When the *ADE2* locus was deleted from the chromosome (*ade2Δ::TRP1*), the rate of colony formation decreased in the parent, *hdf1* and *dnl4* strains, indicating that all are able to perform homologous DSB of the plasmid only when the chromosomal *ADE2* gene is present. Consistent with this, disruption of *RAD52* in *ade2Δ::TRP1* yeast no longer had any effect. In contrast, disruption of either *HDF1* or *DNL4* in *ade2Δ::TRP1* yeast led to a 50-fold reduction in the rate of colony formation, again showing that both are required for NHEJ. Plasmid expression of GST-Dnl4 restored NHEJ in the *dnl4* strain, whereas overexpression of GST-Cdc9 had no effect.

Residual colonies formed by *hdf1* yeast harbour plasmids with imprecise rather than precise junctions^{2,3,5}. Surprisingly, the *ade2Δ::TRP1* parent strain gave colonies with imprecise joins at a

measurable rate, such that $6 \pm 3\%$ of the total colony yield was *ade2* following transformation with *Bgl*II-digested pES16. Moreover, *hdf1* and *dnl4* disruptants did not show the expected increase in the rate of imprecise joining, but rather a consistent and significant decrease relative to the parent strain (Table 2). However, imprecisely joined plasmids recovered from the parent strain all contained only short insertions and deletions at the joined *Bgl*II termini, in essence 'repairing' all but one to a blunt end before ligation (Table 2 and Fig. 4). In contrast, imprecisely joined plasmids recovered from *hdf1* or *dnl4* yeast all showed large deletions with *Bgl*II-digested pES16 (≥ 0.1 kb; Table 2), previously described as the 'error-prone' pathway^{3,5}. We note that two-thirds of plasmids recovered from *hdf1* or *dnl4* yeast did not have large deletions after double digestion of pES16 with *Bgl*II/*Kpn*I, but the rate of imprecise joining at these incompatible ends was still reduced 10- to 20-fold compared with the parent strain (data not shown). Thus it was the loss of the dominant precise joining pathway as well as a microdeletion/insertion end-processing pathway that led to the increased use of the error-prone macrodeletion pathway in *hdf1* and *dnl4* yeast.

We conclude that Dnl4 is the ligase that normally functions downstream of the Ku proteins Hdf1 and Ku80 in both precise and end-processed NHEJ reactions, and that it is neither necessary for (this work) nor sufficient to complement Cdc9-mediated replication, repair and recombination processes¹¹⁻¹⁴. Thus, yeast has two functionally separate pathways for DNA ligation. Cdc9 is used in

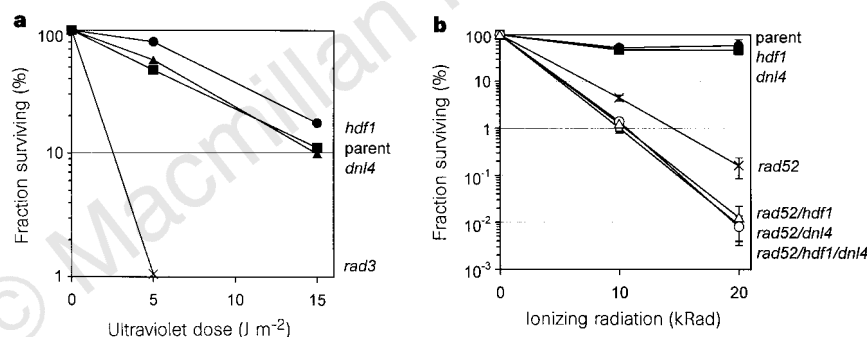


Figure 3 Dnl4 is required for NHEJ but not for excision or homologous DNA repair. **a**, The YW112 parent strain (■), its derivatives with the *hdf1* (●) and *dnl4* (▲) mutations, and the yeast strain X36B-3C (cross) were plated to YPAD medium and exposed to ultraviolet light at the doses indicated. Colonies were counted after two days at 30 °C in order to calculate the surviving fraction. The same result was achieved in two independent experiments. X36B-3C is a UV-sensitive control strain (genotype *MATa*, *rad3-2*, *ade2-1*, *ade4*). **b**, Haploid yeasts were grown to saturation in YPAD and then irradiated in water with a ^{60}Co source with the total

doses indicated. Serial dilutions were plated to YPAD and the colonies counted after 3 days at 30 °C in order to calculate the surviving fraction. Strains were the YW112 parent (■), its derivatives with the *hdf1* (●) and *dnl4* (▲) mutations, YW130 (*rad52*, cross), YW132 (*rad52/hdf1*, □), YW136 (*rad52/dnl4*, ○), and YW140 (*rad52/hdf1/dnl4*, △). Results are the mean $\pm$ standard deviation of at least two independent experiments for YW112 and its derivatives, and at least three independent experiments for YW130-140.

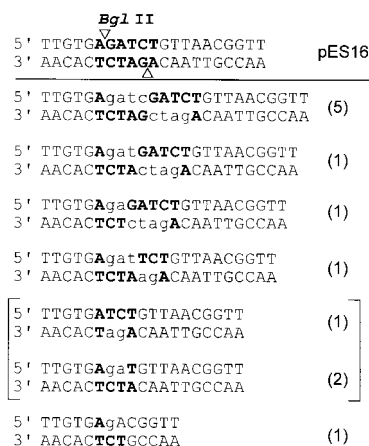


Figure 4 Imprecise NHEJ events in *HDF1/DNL4* yeast show microdeletion and end-filling. At the top is the sequence surrounding the *Bgl*II site of pES16, with cut sites indicated by triangles. Below are the sequences of imprecisely joined plasmids, as recovered from 12 independent *ade2* colonies following transformation of YW112 (*HDF1/DNL4*) with *Bgl*II-digested pES16. Nucleotides likely to have been derived from the original plasmid are shown in upper case, with the original *Bgl*II nucleotides highlighted in bold (note that some have been deleted). *De novo* nucleotides likely to have been derived from templated polymerization of the *Bgl*II 5' overhangs are shown in lower case. The number of times the sequence was recovered is indicated to the right; the bracket surrounds two joins that are themselves symmetric occurrences of the same joining mechanism.

processes where the terminal reaction is ligation of a single-stranded nick in double-stranded DNA. In contrast, Dnl4 is used for double-stranded ligation, an activity that by analogy we predict for mammalian DNA ligase IV. In support of this, the accompanying paper²⁰ shows that the XRCC4 protein, which causes NHEJ and V(D)J recombination defects when absent, interacts functionally with mammalian DNA ligase IV. □

Methods

Multiple sequence alignment. First, homology blocks were identified with the MACAW program²¹ that aligned all four proteins to clear or probable statistical significance, which resulted in the alignment of most of the catalytic domain. Next, statistically significant blocks of the ligase I and ligase IV human/yeast homologous pairs were aligned individually, followed when possible by the best logical alignment of the intervening residues. Not all regions of the proteins can be meaningfully aligned.

Yeast strain construction. Except at the loci indicated, all yeast strains used in the plasmid transformation assay are isogenic with the haploid *ade2-101* parent strain YW104 (also called YPH499, genotype *MATa*, *ura3-52*, *trp1Δ-63*, *his3Δ-200*, *leu2Δ-1*, *lys2-801*, *ade2-101*). The *ade2Δ::TRP1* variant parent strain, YW112, was derived from YW104 by replacing the entire chromosomal *ADE2* sequence with *TRP1*. Disruptions of *HDF1*, *DNL4* and *RAD52* were made in YW104 and YW112 by complete replacement of the relevant coding sequence with the *HIS3* gene, followed by PCR verification²². Strain designations are as follows. YW118: *ade2-101*, *hdf1Δ::HIS3*; YW119: *ade2-101*, *dnl4Δ::HIS3*; YW120: *ade2-101*, *rad52Δ::HIS3*; YW113: *ade2Δ::TRP1*, *hdf1Δ::HIS3*; YW114: *ade2Δ::TRP1*, *dnl4Δ::HIS3*; YW115: *ade2Δ::TRP1*, *rad52Δ::HIS3*. The multiply mutated strains used in the radiation-sensitivity experiments were generated by mating and sporulation using YW113, YW114 and CGY2897 (genotype *MATα*, *ura3-52*, *trp1Δ-63*, *lys2Δ-201*, *his3Δ-200*, *leu2*, *ade2*, *rad1Δ::ADE2*, *rad52::LEU2*). Strain designations are as follows (all strains are additionally *RAD1*, *ura3-52*, *trp1Δ-63*, *his3Δ-200*, *lys2*, *leu2*, *ade2Δ::TRP1*). YW130: *MATα*, *rad52::LEU2*; YW132: *MATα*, *rad52::LEU2*, *hdf1Δ::HIS3*; YW136: *MATa*, *rad52::LEU2*, *dnl4Δ::HIS3*; YW140: *MATα*, *rad52::LEU2*, *hdf1Δ::HIS3*, *dnl4Δ::HIS3*. The presence of both the *HDF1* and *DNL4* deletions in YW140 was verified by PCR of the mutant alleles.

Plasmid construction. GST expression plasmids contained the following PCR-generated fragments in order in the plasmid backbone pRS316 (*URA3*, *CEN/ARS*): the *ADH1* promoter from plasmid pEG202 (ref. 23); the GST coding sequence from plasmid pGEX-3X (Pharmacia); the *CDC9* (pTW268), *DNL4* (pTW269) or *HDF1* (pTW271) coding sequences; and the 1 kb *HDF1* 3' untranslated sequence. pES16 is pRS316 with a PCR fragment corresponding to the 2.5-kb *ADE2* genomic locus inserted at the *EcoRI* site in the polylinker. pES26 was derived from pES16 by addition of a polyterminator oligonucleotide (with *BglII* ends) into the *BglII* site of the *ADE2* gene.

Immunoprecipitation and adenylation. Yeast cells from 25-ml log-phase cultures were lysed by sonication on ice into 1 ml of IP buffer (25 mM HEPES/KOH, pH 7.9, 150 mM KCl, 10 mM MgCl₂, 1 mM DTT, 0.1% NP-40, 1 mM PMSF, 2.5 μg ml⁻¹ bestatin, 2 μg ml⁻¹ aprotinin, 1 μg ml⁻¹ pepstatin A, and 1 μg ml⁻¹ leupeptin). Following immunoprecipitation with 40 μg GST(B-14) antibody (Santa Cruz), beads were washed with IP buffer, divided into two aliquots, and each preincubated in 1 ml ligase buffer (60 mM Tris, pH 8.0, 10 mM MgCl₂, 2 mM DTT, plus protease inhibitors) with or without 0.5 mM sodium pyrophosphate for 5 min at room temperature. Beads were next washed with ligase buffer, and 30 μCi of [α-³²P]ATP mixed with the minimal

residual volume for 30 min at room temperature. After 6% SDS-PAGE, proteins were transferred to nitrocellulose and a Phosphorimager screen exposure and western blot with the GST(B-14) antibody obtained sequentially from the same membrane.

Plasmid transformation assay. Yeast cells were grown to log phase and transformed with 2–5 μg linearized test plasmid by simple lithium acetate transformation without the addition of carrier DNA²⁴. An identical parallel transformation was always done with undigested pES16 or pES26 for the purpose of normalization, so that results for a given strain/digested plasmid combination are expressed as the following ratio of colony counts: [(test strain with digested plasmid)/(test strain with undigested plasmid)]/[reference strain with digested plasmid]/[reference strain with undigested plasmid]].

Received 19 December 1996; accepted 23 May 1997.

- Feldmann, H. & Winnacker, E. L. A putative homologue of the human autoantigen Ku from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **268**, 12895–12900 (1993).
- Milne, G. T., Jin, S., Shannon, K. B. & Weaver, D. T. Mutations in two Ku homologs define a DNA end-joining repair pathway in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 4189–4198 (1996).
- Boulton, S. J. & Jackson, S. P. *S. cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093–5103 (1996).
- Moore, J. K. & Haber, J. A. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *S. cerevisiae*. *Mol. Cell. Biol.* **16**, 2164–2173 (1996).
- Boulton, S. J. & Jackson, S. P. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res.* **24**, 4639–4648 (1996).
- Lieber, M. R. Immunoglobulin diversity: rearranging by cutting and repairing. *Curr. Biol.* **6**, 134–136 (1996).
- Weaver, D., Boubnov, N., Wills, Z., Hall, K. & Staunton, J. V(D)J recombination: double-strand break repair gene products used in the joining mechanism. *Ann. N.Y. Acad. Sci.* **764**, 99–111 (1995).
- Jeggo, P. A., Taccioli, G. E. & Jackson, S. P. Ménage à trois: double strand break repair, V(D)J recombination and DNA-PK. *Bioessays* **17**, 949–957 (1995).
- Lindahl, T. Recognition and processing of damaged DNA. *J. Cell. Sci. (suppl.)* **19**, 73–77 (1995).
- Lindahl, T. & Barnes, D. E. Mammalian DNA ligases. *Annu. Rev. Biochem.* **61**, 251–281 (1992).
- Wilcox, D. R. & Prakash, L. Incision and postincision steps of pyrimidine dimer removal in excision-defective mutants of *Saccharomyces cerevisiae*. *J. Bacteriol.* **148**, 618–623 (1981).
- Tomkinson, A. E., Tappe, N. J. & Friedberg, E. C. DNA ligase I from *Saccharomyces cerevisiae*: physical and biochemical characterization of the *CDC9* gene product. *Biochemistry* **31**, 11762–11771 (1992).
- Moore, C. W. Ligase-deficient yeast cells exhibit defective DNA rejoining and enhanced gamma ray sensitivity. *J. Bacteriol.* **150**, 1227–1233 (1982).
- Barker, D. G., White, J. H. & Johnston, L. H. The nucleotide sequence of the DNA ligase gene (*CDC9*) from *Saccharomyces cerevisiae*: a gene which is cell-cycle regulated and induced in response to DNA damage. *Nucleic Acids Res.* **13**, 8323–8337 (1985).
- Hsieh, C. L., Arlett, C. F. & Lieber, M. R. V(D)J recombination in ataxia telangiectasia, Bloom's syndrome, and a DNA ligase I-associated immunodeficiency disorder. *J. Biol. Chem.* **268**, 20105–20109 (1993).
- Petrini, J., Donovan, J. W., Dimare, C. & Weaver, D. T. Normal V(D)J coding junction formation in DNA ligase I deficiency syndromes. *J. Immunol.* **152**, 176–183 (1994).
- Wei, Y. F. *et al.* Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol. Cell. Biol.* **15**, 3206–3216 (1995).
- Robins, P. & Lindahl, T. DNA ligase IV from HeLa cell nuclei. *J. Biol. Chem.* **271**, 24257–24261 (1996).
- Siede, W., Friedl, A. A., Dianova, I., Eckardt-Schupp, F. & Friedberg, E. C. The *Saccharomyces cerevisiae* Ku autoantigen homologue affects radiosensitivity only in the absence of homologous recombination. *Genetics* **142**, 91–102 (1996).
- Grawunder, U. *et al.* Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**, 492–495 (1997).
- Schuler, G. D., Altschul, S. F. & Lipman, D. J. A workbench for multiple alignment construction and analysis. *Proteins: Struct. Funct. Genet.* **9**, 180–190 (1991).
- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroite, F. & Cullin, C. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**, 3329–3330 (1993).
- Gyuris, J., Golemis, E., Chertkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791–803 (1993).
- Ito, H., Fukuda, Y., Murata, K. & Kimura, A. Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168 (1983).

Acknowledgements. We thank M. Johnston for reagents for yeast-gene disruption, and for helpful discussion on the design of these experiments. T.E.W. is a Howard Hughes Medical Institute physician postdoctoral fellow. This research is supported by NIH grants to M.R.L. M.R.L. is a Leukemia Society of America Scholar.

Correspondence and requests for materials should be addressed to M.R.L. (e-mail: lieber_m@froggy.hsc.usc.edu).

Light fare

This serving of laser equipment will have to sustain readers until next year, offering a variety of lasers, software for laser measurement, a compact optical isolator for diode lasers, as well as mirror, filter and optics accessories.

Isoplot

From Uniscan

New software and image displays for the company's laser triangulation profiler

Uniscan offers analytical spatial profiling systems that include a laser triangulation-based, three-dimensional surface profiler. The three-dimensional profiler is available in the UK through AG Electro-Optics. Uniscan has announced the availability of its Isoplot Scientific graphics software for the OSP-100 profiler. According to the manufacturer, Isoplot produces high-quality, three-dimensional images of data matrices in excess of 200,000 data points. The three-dimensional views can be manipulated to show the desired features of the data and hard copy produced from the standard Windows environment. The software is Windows95 compatible and a 100 MHz Pentium computer is required with a minimum of 16 MB of RAM. Isoplot is capable of producing three-dimensional plots where the data is shaded according to the projected light reflecting from the surface. The z axis may also be amplitude colour shaded in a user-defined palette.

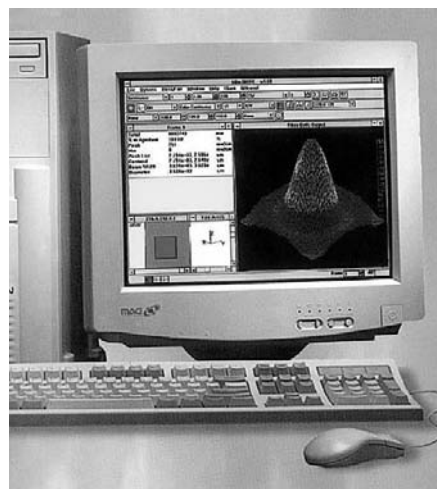
Reader Enquiry No. 100

LBA-300PC software

From Laser Lines

Version 1.2 of this software runs under the 32-bit mode of Windows95, which speeds beam profile access and processing

Featured in this version of the software is beam convolution, which enables the user to remove noise from both the beam view and beam calculations. Primary features of the LBA-300PC include the ability to oper-



Laser Lines' LBA-300PC software works with all RS170 and CCIR cameras for laser beam testing.

ate with all RS170 and CCIR cameras for laser beam diagnostics of visible and near infrared lasers. The program runs under Windows 3.1x or Windows95 on a Pentium computer with a new high-speed PCI bus. According to the manufacturer, the most important feature of the system is an automatic calibration technique that sets the baseline of the camera precisely at zero and saves negative numbers for use in precise beam calculations. This feature allows the camera system to make second-moment, beam-width measurements. The resolution is selectable from 512×480 down to 16×15 , in steps of two. This option to change resolution allows a user to choose the needed resolution and make the trade-off between resolution and operating speed, as well as disk file space. With smaller resolution, less memory is required to store data frames. Lower resolution frames can be obtained either by zooming at single-pixel resolution to a small area, for example 512×480 , 256×240 , 128×120 , 64×60 , 32×30 or 16×15 ; or, by keeping the zoom to cover the entire sensor, reading every first, second, fourth, eighth or sixteenth pixel and line. Thus the user can obtain a low-resolution view of a large beam if it has low structure.

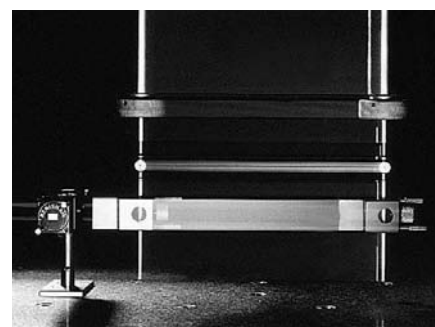
Reader Enquiry No. 101

Dynamic Software

From Renishaw

For the measurement of critical CNC machine performance variables using both time- and distance-based capture and analysis

Used in conjunction with the company's laser measurement system, version 3.0 of the software has a stated sampling rate of up to 5 kHz with a resolution of 1 nm . Data can be captured at up to 1 m s^{-1} , which is said to be 30 per cent faster than previously available, with no additional hardware being required. Applications include condition monitoring of machines and systems involving the measurement of displacement, velocity and acceleration. The system can be used on machine tools, coordinate measuring machines, hydraulic and pneumatic systems, positioning stages, optical stages and PCB drilling machines. Using the package, it is possible to measure displacement, angle, straightness, velocity and acceleration with respect to time. The software operates under Windows (3.1, 3.11 and 95) providing support for Windows printing, cut, copy and paste functions and VGA screen support. The installer sets up the package with start-up icons in a pro-



Dynamic software version 3.0 from Renishaw can be used in conjunction with the company's laser measurement system.

gram group. It is also available in Japanese, Chinese and Korean language versions.

Reader Enquiry No. 102

Lasers

Compact optical isolator for diode lasers

From New Focus

Compact Faraday optical isolators designed specifically for diode lasers

These isolators are designed to provide 35-dB peak isolation and isolate over more than 20 nm without requiring any adjustments. A large elliptical aperture ($3 \text{ mm} \times 8 \text{ mm}$) is said to eliminate the need for beam-shaping optics between the laser and the isolator. The isolator package was designed for easy mounting and compatibility with all standard mounting hardware. A half waveplate is included at the output of each isolator, giving the user complete flexibility in polarization orientation. The company's magnetic field design results in a compact $3.8 \text{ cm} \times 3.8 \text{ cm} \times 8 \text{ cm}$ package and low magnetic field leakage, says New Focus.

Reader Enquiry No. 103

OPC-D015-mmm-HB

From Opto Power Corporation

A high brightness, high-power diode laser module that delivers from 13 to 15 W of continuous-wave (CW) power

Available at operating wavelengths ranging from 790 nm to 980 nm, this compact laser is offered with a wide selection of internal focusing optics that produce working distances as long as 30 mm and spot sizes as small as $130 \text{ }\mu\text{m}$. This module is environmentally sealed and comes with an on-board monitor photodiode and a $10 \text{ k}\Omega$ at $25 \text{ }^\circ\text{C}$ thermistor. According to the manufacturer, this laser provides an order of magnitude improvement in brightness compared to fibre-coupled diode laser arrays and

increases the usefulness of diode lasers for numerous direct thermal and illumination applications, particularly those in industrial, medical and graphics markets that require high brightness and high power.

Reader Enquiry No. 104

Verdi visible CW laser

From Coherent

The company has released a 2-W version of the CW green (532 nm) diode-pumped Nd:YVO₄ laser

The laser uses the same ring laser cavity as the existing 5-W Verdi, which is stated to result in a single-frequency output with exceptionally low noise. In the 2-W version, a single, fibre-delivered diode bar is used as the pump source, rather than the two diode bars used in the 5-W version. Both the 2-W and 5-W versions are stated to be highly efficient and compact, requiring only standard 110/220 V single-phase power and no external cooling. According to the manufacturer, this solid-state technology is well suited for both scientific and industrial applications, including spectroscopy, reprographics, semiconductor inspection, holography, particle scattering, disk-zone texturing and medical applications.

Reader Enquiry No. 105

GreenLine 1000

From Edinburgh Instruments

An all-solid-state, high-power, single-frequency laser with the diode-pumped Nd:YLF laser

The company has broken the 1-W barrier for devices emitting stable, continuous-wave, single-frequency radiation at 523.5 nm. Large-scale holography and pumping of IR, CW optical-parametric oscillators are just two of the applications that are addressed with this product, says the manufacturer. The company expects to expand its experience in all-solid-state, single-frequency lasers into other spectral regions with the development of daughter products and variants of the GreenLine 1000.

Reader Enquiry No. 106

ILT 6800

From Ion Laser Technology

An argon ion laser that offers 1,000 MW power

Available in the UK from AG Electro-Optics, the 6800 series provides over 1 W of power in a multi-line configuration. Single-line or tunable output from 457 nm to 514 nm are also available. The transverse beam mode is TEM₀₀, with higher order modes available on multi-line models. Beam noise is less than one per cent from DC to 2 MHz, and pointing stability is better than 10 microradians after warm-up. According to the manufacturer, the 6800

offers versatility in applications by virtue of the precision light feedback mechanism, which monitors the laser output and amplitude stability, and through the use of a robust power supply with power-factor-correction circuitry. The water jacket for tube cooling is said to eliminate minute vibrations caused by acoustic coupling from cooling fans, and extends the useful life of the tube. The resonator design is said to provide a wide range of fine adjustment of beam parameters with improved stability. Optional features include a remote control with standby interlock, a quick-change feature from single-line to multi-line operation and a digital interface for remote diagnostics or host-based control. The 6800 series is stated to meet international regulatory standards including the CE mark.

Reader Enquiry No. 107

Laser accessories

LDC-3700 controller

From ILX Lightwave

Precision laser diode controllers representing the next generation in their family of precision laser diode controllers

New features include a new processor platform that allows a more robust, high-speed GPIB interface for demanding remote-control operations. Each of the LDC-3700 precision current controller sources provides selectable constant-current power operation, an analog modulation capability, as well as triggering functions for complete system solutions. These new instruments have been redesigned for an expanded dual-current range and are available with maximum currents from 100 mA to 4,000 mA for driving a variety of laser diodes. The TEC controllers offer 16 W of control power with a choice of temperature sensors for improved precision temperature control.

Reader Enquiry No. 108

Flipper mount mirror

From New Focus

This motorized mount allows the automation of multiple experiments

The company has added a motor to their Flipper mirror mount, which now lets users repeatedly flip a mirror in and out of a beam path remotely. A single 9-V battery provides hundreds of flips, and the unit can be controlled through a TTL level input. Manual alignment screws are provided for adjusting both the tip and the tilt of the mirror. These mirrors should prove useful for applications where users need to share laboratory equipment. Also available is the stackable Flip-Up optics mount. The model 9897 Flip-Up mount has been designed to allow easy addition and removal of filters and lenses from experiments. The mount can be easily



Also available from New Focus are Flip-Up optic mounts for easy changing of lenses and filters.

stacked to change filter values quickly. Both thin and thick optics are held firmly using either a threaded Delrin insert or a standard set screw.

Reader Enquiry No. 109

UVS laser accessory

From Laser Sciences

An accessory that has been developed to provide economical ultraviolet laser sources

When coupled to the company's pulsed nitrogen laser and laser dye module, this accessory produces wavelength selectable light from 220 nm to 320 nm. This is a compact unit, which is said to be suitable for ultraviolet laser-induced fluorescence applications in clinical, pharmaceutical, environmental and process control work.

Reader Enquiry No. 110

Internet resource

Located on the World Wide Web, Tate Web offers information and links to optics and photonics resources: <www.tateweb.com>. □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

High Quality **ADR Laser Pointers** on special offer 10% off from £21.60 (VAT and P&P exclusive)

Impressive High-tech Gift

Appreciated by all (Can be personalised for quantity corporate gift orders, of which prices are drastically reduced) **Powerful Presentational Tool** Giving you Complete Command

Contact NOW for a free leaflet Official and Overseas orders welcome!

Adaptive Dynamics Research Ltd.
The Innovation Centre
University of Wales Swansea
Singleton Park,
Swansea SA2 8PP, UK
ULR:- <http://www.demon.co.uk/adrlh>



Tel: 01792 295012 (UK),
Fax: 01792-208953 (UK),
E-mail: adrhl@adrhl.demon.co.uk